ANALYSIS OF THE MOVEMENT OF *SACCHAROMYCES CEREVISIAE* MISMATCH REPAIR PROTEINS ON DNA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Aaron J. Plys

August 2011

ANALYSIS OF THE MOVEMENT OF *SACCHAROMYCES CEREVISIAE* MISMATCH REPAIR PROTEINS ON DNA

Aaron J. Plys, Ph. D.

Cornell University 2011

Replication errors that escape DNA polymerase proof-reading activity are efficiently recognized and repaired by conserved DNA mismatch repair factors. The overall result is a drastic reduction in deletion mutations. The mechanistic details of how mismatch repair proteins execute mismatch removal have not been elucidated. The aim of my thesis is to better understand how mismatch repair factors interact with DNA in order to identify mismatch sites.

My work reveals that the mismatch repair complex, MLH1-PMS1, has unique DNA diffusion characteristics facilitated by structural features of the two subunits. Through bulk assays and total internal reflectance fluorescence microscopy (TIRFM), I found that MLH1-PMS1 could independently bind DNA and rapidly diffuse using the thermal energy of the system. Furthermore, MLH1-PMS1 was shown to be the first passively diffusing protein that could bypass stationary nucleosomes. In contrast, the DNA diffusion activity of the mismatch recognition complex MSH2-MSH6 was blocked by nucleosomes. The timing and nature of mismatch repair is linked with replication and is thus proposed that the differences seen for the two complexes have important implications for repair in the context of the chromatin state directly at the replication fork.

Each subunit of the MLH1-PMS1 complex is composed of two defined globular domains connected by an unstructured linker arm. The linker arms of the complex are

proposed to facilitate topological DNA binding and diffusion along DNA in a hopping/stepping mechanism. I found that TEV protease cleavage within the linker arms of MLH1-PMS1 disrupted DNA binding and mismatch repair *in vitro* and *in vivo*. Using a genetic mismatch repair assay I found that shortening of the linker arms in MLH1 had a drastic effect on function whereas similar changes in PMS1 had little or no effect. Purified truncated complexes were able to interact with DNA and form ternary complexes with MSH2-MSH6 at a mismatch. Future studies should focus on the diffusion characteristic for these complexes. Together, my work has important implications for understanding how mismatch repair proteins can rapidly identify their targets in a chromatin landscape.

BIOGRAPHICAL SKETCH

Aaron Jacob Plys was born on a typical spring day in Minnesota, during a snow storm on March 23, 1983. He was the first of four siblings raised primarily by his mother and her family. Most of his childhood was spent outdoors playing sports, fishing and biking. If he was not playing sports himself, then he was constantly following his favorite Minnesota sports teams (Vikings, Twins, Gophers, Northstars/Wild) on the TV, radio, newspaper, etc. In high school he excelled in the classroom and also on the varsity football and baseball teams. After graduating from high school in 2001, he attended the University of Minnesota to pursue a degree in Biochemistry. 2001 turned out to be the best year to attend the University of Minnesota if you were a huge hockey fan, like Aaron, because the Gophers won the first of back-to-back national championships that year. There were plenty of ups and downs in college, but thankfully he was able to right the ship in his last two years to finish with a double-major in Biochemistry and Genetics in 2005. During his senior year of college he joined the cell biology lab of Dr. Ross Johnson where he studied fluorescently tagged gap junction proteins. It was at this time while doing hands-on novel experiments that Aaron decided to pursue a career in research. Amazingly, he was accepted into the Biochemistry, Molecular and Cell biology doctoral program at Cornell University in 2005. During his time at Cornell he joined the laboratory of Dr. Eric Alani to study the biochemical and genetic properties of DNA mismatch repair proteins in the yeast *Saccharomyces cerevisiae*. At Cornell, is also where he met his classmate and future wife Gizem Rizki. They both plan on moving to Cyprus after graduation to start their lives together.

ACKNOWLEDGMENTS

I would like to thank all of the members of the Alani lab, as they have provided me with extremely helpful advice and support during my time in the lab. Special thanks to Dr. K.T. Nishant for his contributions to Chapter 4 and his help in understanding meiotic recombination. The guidance and biochemical expertise of Dr. Jennifer Surtees was invaluable in my success as a graduate student and as a biochemist. I am greatly indebted to my advisor Dr. Eric Alani for everything that he has done for me during my time in the lab. He was always there for encouragement or advice when it was needed and is exemplary in the way that he manages and guides students. During the time in his laboratory I have learned how to properly conduct and present scientific research and this is a reflection of his guidance.

Much of the work done in Chapter 2 could not have been done without the excellent collaboration with Dr. Eric Greene and Jason Gorman at Columbia University. They helped show me the importance of working with other laboratories to accomplish scientific goals and were exceptionally accommodating when I traveled to New York City to learn their techniques. I would also like to thank my committee members Dr. Bob Weiss and Dr. Ruth Collins for their time and efforts in learning about my projects and the helpful advice that they gave to me.

The work presented here was partially supported by a Diversity Fellowship and two conference travel grants from the Cornell graduate school, several teaching assistantships, an ASBMB travel award, and the Alani lab NIH grant GM53085.

Last, but not least, I would like to thank my friends and family for their support and to let them know that my accomplishments would not have been possible without them. All of

my friends at Cornell made it a fun place to work and they allowed me to enjoy my time spent in Ithaca. Finally, my soon to be wife, Gizem, was always there to listen to my frustrations and offer me encouragement when things were not going well and even more importantly she has been there to share in my joys inside of lab and out.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH..... Page iii

ACKNOWLEDGMENTS.....Page iv

LIST OF FIGURES.....Page vii

LIST OF TABLES.....Page ix

LIST OF ABBREVIATIONS.....Page x

Chapter 1: DNA mismatch repair in mammals.....Page 1

Chapter 2: Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice.....Page 34

Chapter 3: The function of unstructured linker arm domains of MLH1 and PMS1 is consistent with enabling diffusion of the complex along chromatin .....Page 82

Chapter 4: A mutation in the putative MLH3 endonuclease domain confers a defect in both mismatch repair and meiosis in *Saccharomyces cerevisiae*: Published and unpublished results .....Page 119

Chapter 5: Implications for the DNA mismatch repair field and unresolved questions.....Page 142

Appendix .....Page 151

LIST OF FIGURES

Figure 1.1 Mutations seen in mismatch repair defective strains.....Page 4

Figure 1.2 MMR Components.....Page 7

Figure 1.3 Basic Steps in MMR.....Page 9

Figure 1.4 Model for signaling downstream repair.....Page 15

Figure 1.5 How might strand discrimination be accomplished during mismatch repair?.....Page 20

Figure 2.1 Nanofabricated racks of DNA for visualizing 1D diffusion of Mlh1–Pms1.....Page 39

Figure 2.2: DNA-binding activity of fluorescently tagged Mlh1–Pms1.....Page 41

Figure 2.3: Quantitative analysis of Mlh1–Pms1 diffusion .....Page 43

Figure 2.4: End-dependent dissociation of Mlh1–Pms1 from DNA .....Page 46

Figure 2.5. TEV cleavage of Mlh1-Pms1 linker arms disrupts DNA-binding.....Page 49

Figure 2.6: Diffusion of Mlh1–Pms1 and Msh2–Msh6 along nucleosome-bound DNA .....Page 53

Figure 2.7, Supplemental. DNA curtains provide sufficient clearance for free passage of QDs. .....Page 154

Figure 2.8, Supplemental. Salt-dependence of Mlh1-Pms1 1D diffusion .....Page 156

Figure 2.9, Supplemental. Mlh1-Pms1 complexes can occasionally bypass one another while traveling along the same DNA .....Page 157

Figure 3.1. Model for replication-associated mismatch search by MSH2-MSH6 and MLH1-PMS1.....Page 86

Figure 3.2. TEV cleavage of Mlh1-Pms1 linker arms disrupts DNA-binding.....Page 88

Figure 3.3. Schematic diagram of MLH1 and PMS1 linker arm deletion series.....Page 94

Figure 3.4. MLH1-PMS1 linker arm deletions can be expressed and purified .....Page 99

Figure 3.5. *mlh1* and *pms1* linker arm deletions display a range of altered DNA binding affinities .....Page 101

Figure 3.6. *mlh1* and *pms1* linker arm deletion mutants that are able to bind to DNA are also able to form a ternary complex with MSH2-MSH6 and a mismatch .....Page 104

Figure 4.1, Alignment of the conserved DQHA$(X)_2$E$(X)_4$E endonuclease domain between MLH3 homologs.....Page124

Figure 4.2 Epitope-tagged MLH3 and mlh3-D523N are stably expressed and interact with MLH1 .....Page 126

Figure 4.3. Partial purification of epitope-tagged MLH3 with MLH1.....Page 128

Figure 4.4. Partially purified MLH1-HA-MLH3 does not contain a detectable endonuclease activity .....Page 130

Figure 5.1 Speculative model for strand discrimination signal search by DNA mismatch repair factors after mismatch recognition .....Page 144

Figure 5.2, Supplemental. Illustration of complexes containing engineered N-terminal swaps of MLH1-PMS1 .....Page 160

LIST OF TABLES

Table 2.1, Supplemental. Summary of Mlh1-Pms1 behavior from videos.....Page 158

Table 3.1. List of strains and plasmids used in this study.....Page 90

Table 3.2. TEV protease cleavage of *MLH1* confers a mutator phenotype *in vivo* .....Page 92

Table 3.3. *mlh1* and *pms1* linker arm deletions confer differential mutator phenotypes.....Page 96

Table 5.1. Swapping MLH1-PMS1 N-terminal domains results in non-functional complexes displaying high mutator phenotypes .....Page 162

LIST OF ABBREVIATIONS

MMR – Mismatch Repair

DNA – Deoxyribonucleic acid

Indel – insertion/deletion mutation

DSB – Double-strand break

dHJ – Double Holliday junction

ATP – Adenosine triphosphate

ADP – Adenosine diphosphate

bp – base pair

HNPCC – Hereditary Non-Polyposis Colorectal Cancer

E. coli – Escherichia coli

S. cerevisiae – Saccharomyces cerevisiae

TIRFM – Total internal reflectance fluorescence microscopy

QD – Quantum dot

HA – Hemagglutinin

DIG – digoxigenin

FITC – Fluorescein isothiocyanate

1D – One-dimensional

Δ – deletion

wt- Wildtype

NMR – Nuclear magnetic resonance

PCR – Polymerase chain reaction

DTT- Dithiothreitol

EDTA - Ethylenediaminetetraacetic acid

SDS – Sodium dodecyl sulfate

SDS-PAGE – SDS-Polyacrylamide gel electrophoresis

TBE – Tris, boric acid, EDTA

EMSA – Electromobility shift assay

Etbr – Ethidium bromide

PBS – Phosphate buffered saline

PEG – Polyethylene glycol

PMSF – Phenylmethanesulfonylfluoride

IB – Immunoblot

NTD – N-terminal domain

CTD – C-terminal domain

CBD – Chitin binding domain

BSA- Bovine serum albumin

TEV – Tobacco etch virus

CIP – Calf intestinal phosphatase

MSD – Mean squared displacement

C.I. – Confidence interval

S.D. – Standard deviation

RT – Room Temperature

C – Celsius

V – Volts

MW – molecular weight

OD – Optical density

w/v – weight/volume

v/v – volume/volume

min. – minute

sec. – second

kb- kilobase

kDa – kilodalton

Kd – Dissociation constant

fN – Normal force

M – Molar

mM – Millimolar

μM – Micromolar

nM – Nanomolar

L – Liter

ml – Milliliter

μl – Microliter

g – gram

mg – milligram

μg – microgram

mm – millimeter

μm – micrometer

nm – nanometer

Chapter 1

DNA mismatch repair in mammals

Aaron J. Plys and Eric Alani

Department of Molecular Biology and Genetics

Cornell University

Ithaca, NY 14853-2703

This chapter is adapted from a paper that is in press in the 2nd edition of the Encyclopedia of Biological Chemistry:

Plys, Aaron J. and Eric Alani: DNA Mismatch Repair in Mammals. Encyclopedia of Biological Chemistry. 2011. Volume 2, Article No. 315, Pages TBD.

Abstract

Mutations, which are important for the evolution of genomes and impact disease progression, occur at similarly low rates in a variety of organisms. Mammalian genomes, comprised of billions of nucleotides, are replicated with exquisite precision during each cell cycle. The DNA Mismatch Repair (MMR) system serves as a spell checker to correct nucleotide misincorporations that escape detection by DNA polymerases. MMR improves the fidelity of DNA replication so that mutations occur only once in ~10^9 polymerization events. In the absence of MMR mutation rates are increased by ~1000 fold. The vast majority of these mutations are base substitutions and small insertions/deletions that occur primarily in repetitive DNA. In humans mutations in at least four MMR genes, inherited in an autosomal-dominant fashion, have been implicated in hereditary non-polyposis colorectal cancer. Mutations in MMR genes also lead to higher frequency of sporadic tumors in many tissue types through increases in spontaneous mutations. This review provides an overview of the MMR mechanism in mammalian cells, focusing on factors required for mismatch recognition and downstream steps, and interactions with the DNA replication machinery.

Introduction

MMR is an evolutionarily conserved pathway comprising components that work in concert to recognize DNA polymerase-mediated misincorporation errors and facilitate their excision (Buermeyer *et al.* 1999; Kolodner and Marsischky 1999; Harfe and Jinks-Robertson 2000; Marti *et al.* 2002; Kunkel and Erie 2005; Iyer *et al.* 2006; Jiricny 2006; Li 2008). These misincorporations are primarily base-base mismatches (e.g. G-T, C-A) or insertion/deletion loops (indels) that arise from slippages that occur when DNA polymerase is replicating repetitive sequences in the genome (Figure 1.1). The major MMR components, Mut proteins, were identified several decades ago in the bacterium *E. coli* in genetic screens for mutations that increased the cellular mutation rate (Glickman and Radman 1980). Defects in *mut* genes result in an increased rate of base substitution and indel mutations (~1000-fold), as well as increased genetic recombination between divergent DNA sequences and gene amplifications (Modrich and Lahue 1996; Harfe and Jinks-Robertson 2000; Chen *et al.* 2001) Mutant homologs of *mut* genes in mammals also have defects in cell-mediated death and immune system functions (Bellacosa 2001; Martin and Scharff 2002). In humans, the high frequency of indel mutations seen in tumors obtained from a subset of patients suffering from inherited forms of non-polyposis colorectal cancer (2 to 6% of all colon cancers) was a strong hint that these individuals suffered from defects in MMR (Lynch *et al.* 2009). This characteristic resulted in the identification of mutations in at least four MMR genes that are linked to HNPCC (Liu *et al.* 1996). Loss of MMR activity has also been linked to resistance to chemotherapeutic agents (Irving and Hall 2001).

Mammals contain multiple homologs of the *E. coli mut* genes (Buermeyer *et al.* 1999). MutS homologs (MSH family), of which there are five members (MSH2-MSH6), act as

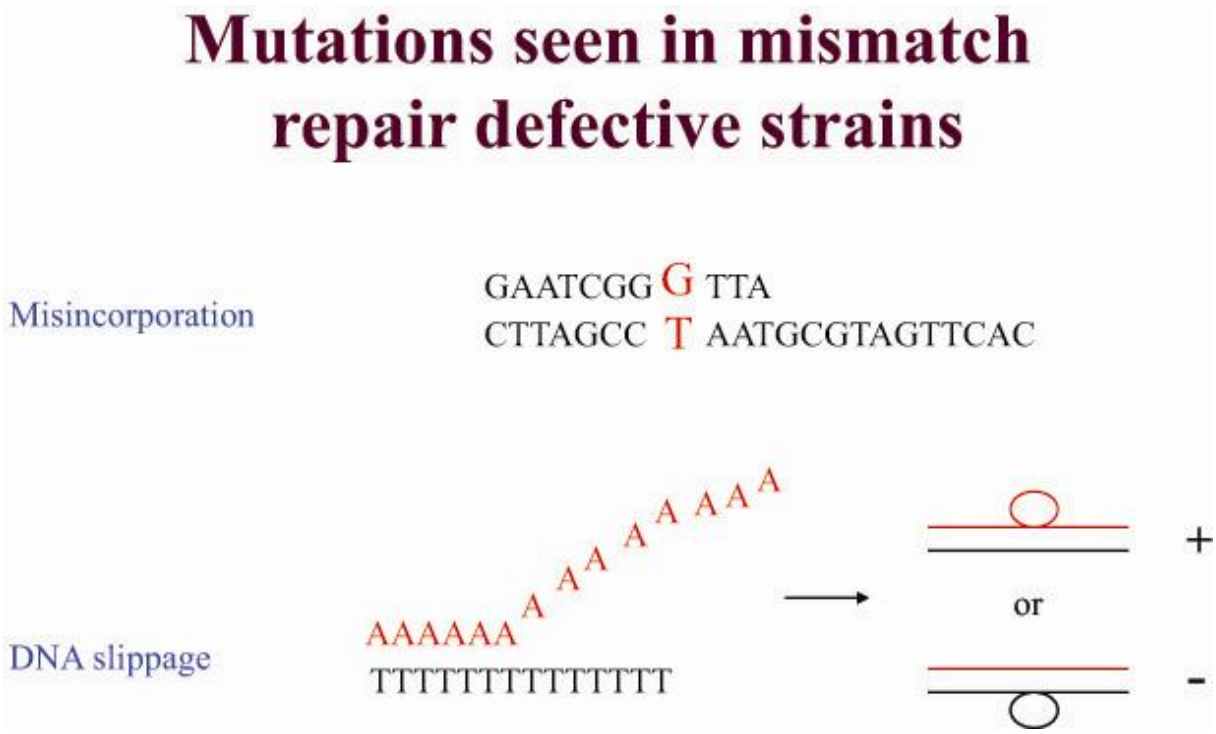


Figure 1.1 Mutations seen in mismatch repair defective strains.
Misincorporation mutations arise from DNA polymerase errors that create non-Watson-Crick base pairs. DNA slippage mutations commonly occur on repetitive DNA elements by DNA strands reannealing out of register. This creates an insertion (+) if the loop is on the newly replicated strand or a deletion (-) if on the template strand.

heterodimers (Figure 1.2). MutS homolog proteins initiate the MMR pathway by recognizing mismatches. The primary complexes are MSH2-MSH6 (MutSα) and MSH2-MSH3 (MutSβ). These complexes have partially overlapping roles in recognizing mismatch substrates. MutSα is primarily involved in repairing base-base mismatches and small indel mutations (Iaccarino *et al.* 1996). MutSβ acts primarily to repair indels that are up to 17 nucleotides in length (Habraken *et al.* 1996; Palombo *et al.* 1996). MSH4 and MSH5 also form a heterodimeric complex that has no apparent role in MMR but acts to promote crossing over in meiosis (Hollingsworth *et al.* 1995). The lower eukaryote baker's yeast contains an MSH1 protein involved in mitochondrial genome stability; however, no mammalian homolog of MSH1 has been identified (Chi and Kolodner 1994). Similarly, there are multiple MutL homologs (MLH family), MLH1, MLH3, PMS1, and PMS2 that function as heterodimers. The function of MutL homologs is to bridge mismatch recognition with the downstream repair steps of MMR. MLH1-PMS2 (MutLα) and MLH1-MLH3 (MutLγ) also have partially overlapping roles in initiating downstream repair from MMR targets. MutLα is the predominant complex in repairing mismatches and indels while MutLγ participates in only a subset of indel repair and also participates in promoting crossing over in meiosis (Wang *et al.* 1999). MLH1 and PMS1 also form a heterodimeric complex; however, the function of this complex in mammals is less clear (Raschle *et al.* 1999). Lastly, an MLH2 homolog has been identified in baker's yeast and has been implicated in the repair of frameshift mutations; however, an equivalent factor has not been identified in mammals (Harfe *et al.* 2000). The downstream components of the mammalian MMR pathway do not belong to the Mut homolog families and are not MMR specific.

A key breakthrough in understanding the mechanistic details of the MMR pathway came from the development of a reconstituted MMR system from purified *E. coli* components

(Lahue *et al.* 1989). The substrate for this assay was a hemi-methylated plasmid containing a single mismatch; *in vitro* repair resulted in excision of the mismatch on the unmethylated strand (see later). This achievement led to a detailed characterization of individual bacterial MMR components, and has served as the basis for characterizing key mammalian MMR factors.

In *E. coli*, MMR is initiated by a MutS homodimer binding to a DNA mismatch (Figure 1.2). Interactions between MutS, MutL, and mismatch DNA result in the removal of the mismatch through an excision mechanism (Grilley *et al.* 1993). Excision is initiated by activation of the MutH endonuclease in steps that require MutS binding to mismatch DNA, MutL interaction with the MutS-mismatch complex, and ATP (Au *et al.* 1992). MutH, a single-strand DNA cleaving enzyme that acts at unmethylated GATC sites (Welsh *et al.* 1987), plays a critical role in a strand discrimination mechanism that removes the mismatch on the newly replicated strand. Strand discrimination is accomplished by a MutH cleavage activity that is inhibited by Dam, a protein that specifically methylates adenine residues at GATC sites (Hattman *et al.* 1978; Brooks *et al.* 1983). Methylation of newly replicated DNA by Dam occurs within a short period after replication fork passage. This post-replication function of Dam establishes a strand discrimination signal because MutH will not cleave methylated GATC sites but will cleave the unmethylated strand of hemi-methylated GATC sites. Thus DNA mismatches that are recognized shortly after DNA replication fork passage are directed for removal of the newly replicated strand due to MutH directed cleavage of hemi-methylated GATC sites (Messer and Noyer-Weidner 1988). MutH incision at hemi-methylated GATC sites can occur up to several thousand base pairs from the mismatch, with repair efficiency decreasing with distance (Lahue *et al.* 1989).

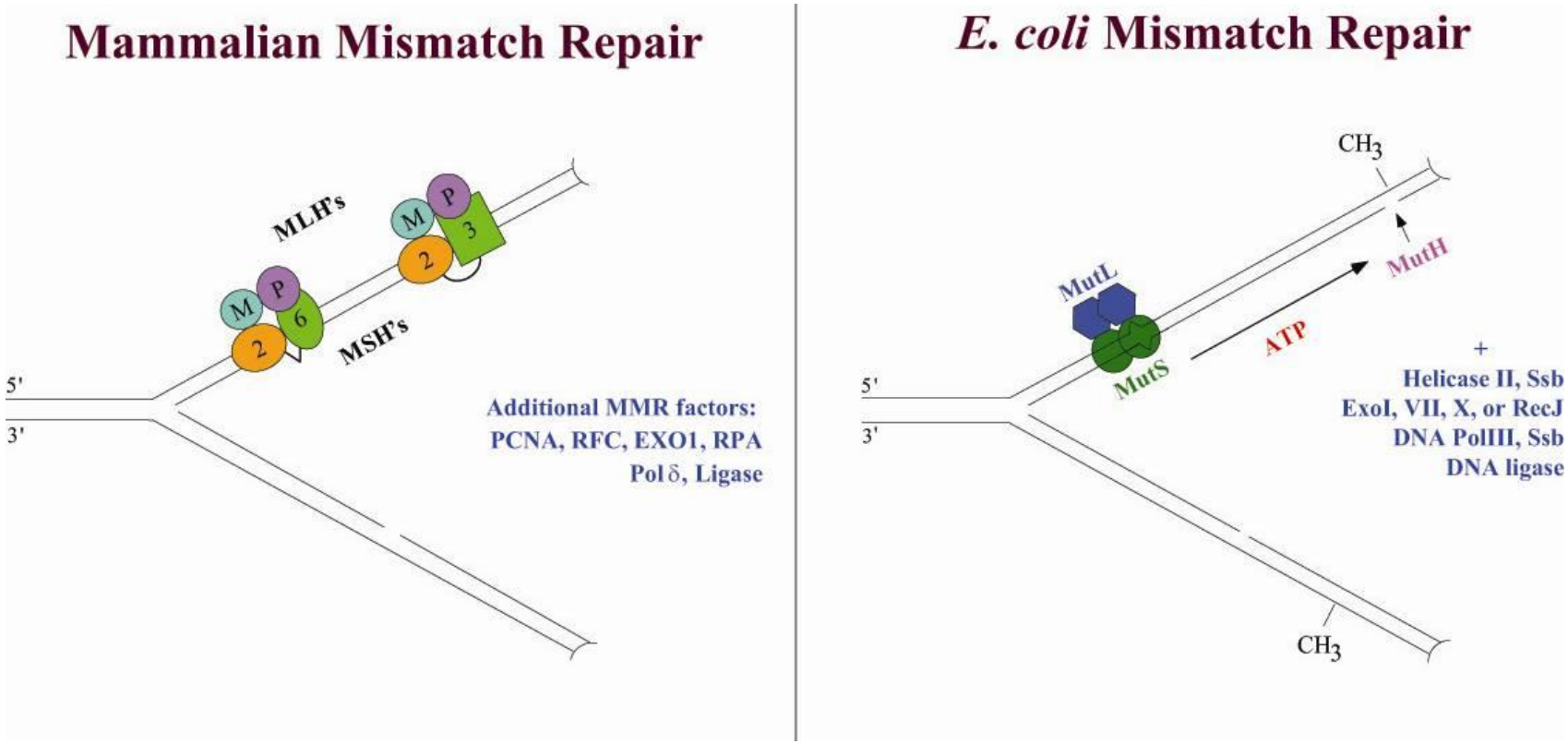


Figure 1.2 MMR Components
Left panel: Mammalian mismatch repair. M = MLH1, P = PMS2, 2 = MSH2, 3 = MSH3, 6 = MSH6. A base-base mismatch is shown by the triangle and an indel created by DNA slippage is indicated by a half circle. Additional factors in mammalian MMR are shown. Right panel: Model replication fork from *E. coli.* A base-base mismatch is indicated by a diamond. Methylation by Dam is indicated by –C

MutH incision creates an entry site for UvrD, a helicase that can unwind duplex DNA beginning at a nick site with the aid of single-stranded DNA binding protein (SSB) (Runyon *et al.* 1990). The MutS-MutL-mismatch complex provides signals to load UvrD and unwind DNA towards the mismatch (Matson and Robertson 2006). The unwound single-stranded DNA containing the nicked end is acted upon by at least four exonucleases (ExoI, ExoVII, ExoX, RecJ) (Burdett *et al.* 2001; Viswanathan *et al.* 2001). These exonucleases display different polarities (5' to 3' or 3' to 5') for excision. Which nuclease acts depends on the orientation of the mismatch relative to which MutH cleavage site is chosen. Thus the MMR system is capable of bi-directional repair (5' or 3' directed excision) and is able to sense the location of the mismatch to direct appropriate excision (Grilley *et al.* 1993). The resulting gap created by excision is filled in by DNA polymerase III with the help of the replication processivity factor β–clamp. DNA ligase closes the nick to complete repair (Figure 1.3).

Mammalian MMR shares aspects of the *E. coli* MMR mechanism with some important differences. Like in the *E. coli* system, the development of a cell-free system to study mismatch repair has been critical in terms of identifying most of the key components (Dzantiev *et al.* 2004; Constantin *et al.* 2005; Zhang *et al.* 2005). The main caveat, discussed here, is the fact that the cell-free system has not been able to address how strand discrimination is set up, other than showing that a pre-existing nick on a mismatch plasmid substrate is sufficient to direct strand-specific repair.

As described earlier, mammalian MSH and MLH factors act as heterodimers to carry out functions analogous to those observed for the *E. coli* proteins. However, there are a least three mechanistic differences between the bacterial and mammalian MMR pathways:

1. A methylation system that distinguishes the template and newly replicated DNA strands does not appear to function in mammals and bona fide MutH homologs appear to be

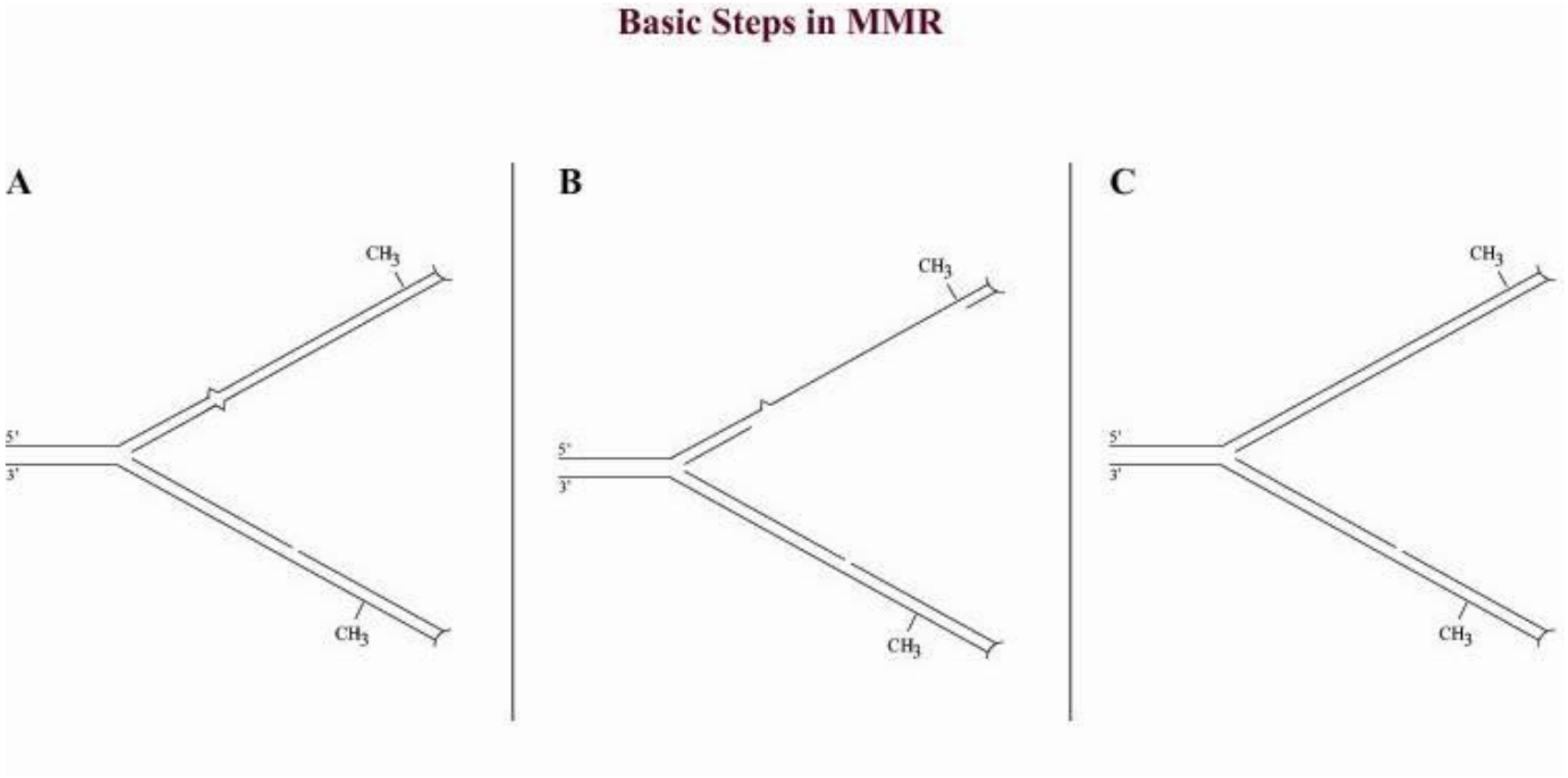


Figure 1.3 Basic Steps in MMR
(A) DNA mismatch is created by a DNA replication error. **(B)** Excision of the mismatch on the newly replicated strand creates a gap. **(C)** Resynthesis of the gap completes repair.

missing in eukaryotes and in most bacteria (Marti *et al.* 2002). Thus, mammalian MMR uses a different mechanism for strand discrimination. Models that outline how the mammalian MMR pathway distinguishes template from daughter strands have been proposed and will be discussed later.

2. A helicase activity analogous to UvrD does not appear to be required in cell-free mammalian MMR reactions (Genschel *et al.* 2002).
3. Exonucleolytic activities that excise the newly replicated strand appear to work differently in mammalian MMR. So far only a single exonuclease, EXO1, has been identified in MMR, and this protein is a 5' to 3' exonuclease that acts primarily on double-stranded DNA (Tran *et al.* 1999). The fact that EXO1 can act on double-stranded DNA may preclude the need for a helicase activity.

Mismatch search and recognition

MutS the mismatch recognition complex:

Crystallographic analysis revealed that the MutS homodimer binds to mismatch DNA in an asymmetric fashion (Lamers *et al.* 2000; Obmolova *et al.* 2000). In the absence of DNA, MutS contains disordered DNA binding domains. Upon mismatch binding these domains undergo extensive local folding to form more ordered structures. Each MutS subunit contains two DNA binding domains, however only one domain of one subunit directly interacts with a mismatched base while a DNA binding domain on the other subunit makes non-specific interactions to stabilize the DNA backbone. The overall structure of MutS resembles a clamp that encircles the DNA backbone (Lamers *et al.* 2000). Studies on the crystal structures of the human MutSα complex have shown similarities with bacterial MutS in terms of domain

organization and DNA interactions (Warren *et al.* 2007). In *E. coli* MutS, mismatch recognition is coordinated by two amino acids, phenylalanine 36 and glutamic acid 38 (Natrajan *et al.* 2003). In mammals, MSH6, but not MSH2 or MSH3, contains these two key residues and is the subunit responsible for direct interactions with the mismatch (Drotschmann *et al.* 2001). In both MutS and MSH6 the mismatch base stacks with the phenylalanine side chain and hydrogen bonds with the glutamic acid side chain. Moreover, the DNA bound to MutS in the crystal structure is in a bent state that is facilitated by these interactions and with non-specific interactions between the other DNA binding domain and the DNA backbone (Wang *et al.* 2003). Recent atomic force microscopy studies have suggested that MutS undergoes a conformational change after binding that results in the formation of unbent DNA containing the mismatch (Wang *et al.* 2003; Tessmer *et al.* 2008). This type of manipulation may allow MutS to sample DNA for mismatches in a manner that increases target recognition specificity.

MSH complexes must identify rare misincorporation errors among a vast excess of undamaged DNA. Biochemical and single-molecule studies have shown that MSH complexes can slide along DNA by facilitated one-dimensional diffusion; such a mechanism is likely to promote efficient identification of targets because the search is confined to scanning along DNA rather than requiring random collisions within the three-dimensional space of the nucleus (Gorman *et al.* 2007). MSH complexes have also been shown to interact with components of the DNA replication machinery (Gu *et al.* 1998; Clark *et al.* 2000; Flores-Rozas *et al.* 2000). This feature, coupled with their ability to diffuse along DNA, suggest that MSH complexes have immediate access to DNA polymerase-mediated misincorporation errors.

ATP binding and hydrolysis by the MSH proteins appear critical for coordinating

mismatch recognition with the commitment to undergo excision repair (Gradia *et al.* 1997). The MSH proteins are members of the ABC transporter adenosine triphosphatase (ATPase) family and weakly hydrolyze adenosine triphosphate (ATP) (Alani *et al.* 1997). Nucleotide binding results in conformational changes in MSH proteins. The nucleotide binding sites of MSH proteins are found at their dimerization interface (Alani *et al.* 2003). MSH2 and MSH6 in MutSα show different affinities for adenine nucleotides. MSH6 has a higher affinity for ATP whereas MSH2 has a higher affinity for ADP (Studamire *et al.* 1998; Martik *et al.* 2004). Nucleotide binding asymmetry is also seen in the *E. coli* MutS homodimer (Lamers *et al.* 2003). Such asymmetries in ATP binding by MSH subunits are thought to be important to induce coordinated conformational changes in MSH-mismatch DNA complexes that signal downstream repair factors. For example in the eukaryotic MutSα (MSH2-MSH6) complex, upon mismatch binding, MSH6 ATP hydrolysis is inhibited and ATP binding by MSH2 is favored (Mazur *et al.* 2006). This ADP --> ATP nucleotide exchange allows MutSα to enter a sliding clamp diffusion mode (Gradia *et al.* 1999). Implications for a MutSα sliding clamp diffusion mechanism in searching out the unknown strand discrimination signal is discussed later (Figure 1.4). ATP is also required for recruitment of MLH proteins to mismatch-MSH complexes to form a ternary complex for the transmission of the mismatch recognition signal to the downstream repair factors (Habraken *et al.* 1998).

MutL bridges mismatch recognition and excision:

As outlined above, the MLH family of proteins plays critical molecular matchmaking roles by signaling MSH-dependent mismatch recognition to downstream repair factors. Crystallographic structures have been reported for individual N- and C- terminal fragments of the *E. coli* MutL protein (Ban and Yang 1998; Ban *et al.* 1999; Guarne *et al.* 2004; Kosinski

et al. 2005). This analysis, described earlier, has shown that the MLH complexes undergo conformational changes that involve primarily the N-terminal domains. The existing structures lack connecting linker arms between the N- and C- terminal domains. Currently, only the N-terminal domains of the yeast and human PMS2 subunits (named PMS1 in yeast) have been solved (Guarne *et al.* 2001; Arana *et al.* 2010).

The N-terminal domain of MutL family proteins contains an ATP binding site that belongs to the GHKL family of ATPases (Dutta and Inouye 2000). This domain in *E. coli* MutL dimerizes in the presence of a nonhydrolyzable ATP analog, AMP-PNP (Ban *et al.* 1999). ATP binding by GHKL family members induces large conformational changes in the proteins and such changes also occur in the ATP bound MutL proteins (Sacho *et al.* 2008). The less-conserved C-terminal domain of MutL family proteins contains a region important for dimerization that does not appear to be modulated by nucleotide or other cofactors (Guarne *et al.* 2004). The unstructured linker arms of MutLα become more ordered upon nucleotide binding and appear to promote interactions between the N-terminal domains of the two subunits (Tran and Liskay 2000).

Like the MSH family proteins, ATP hydrolytic activities appear different among the mammalian MLH subunits and are likely to be functionally relevant for coordinating interactions with downstream repair factors. MLH1 hydrolyzes ATP more proficiently than PMS2 (Hall *et al.* 2002). In the absence of nucleotide the overall structure of full-length MutLα is predicted to be a V-shaped molecule connected at the C-termini that undergoes conformational changes in the linker arms upon nucleotide binding to bring the N-termini together and form a ring like structure (Sacho *et al.* 2008). As previously mentioned, ATP has an apparent role in modulating the interactions of various components of the MMR pathway. MutSα bound to a mismatch is capable of interacting with MutLα in the presence of ATP to

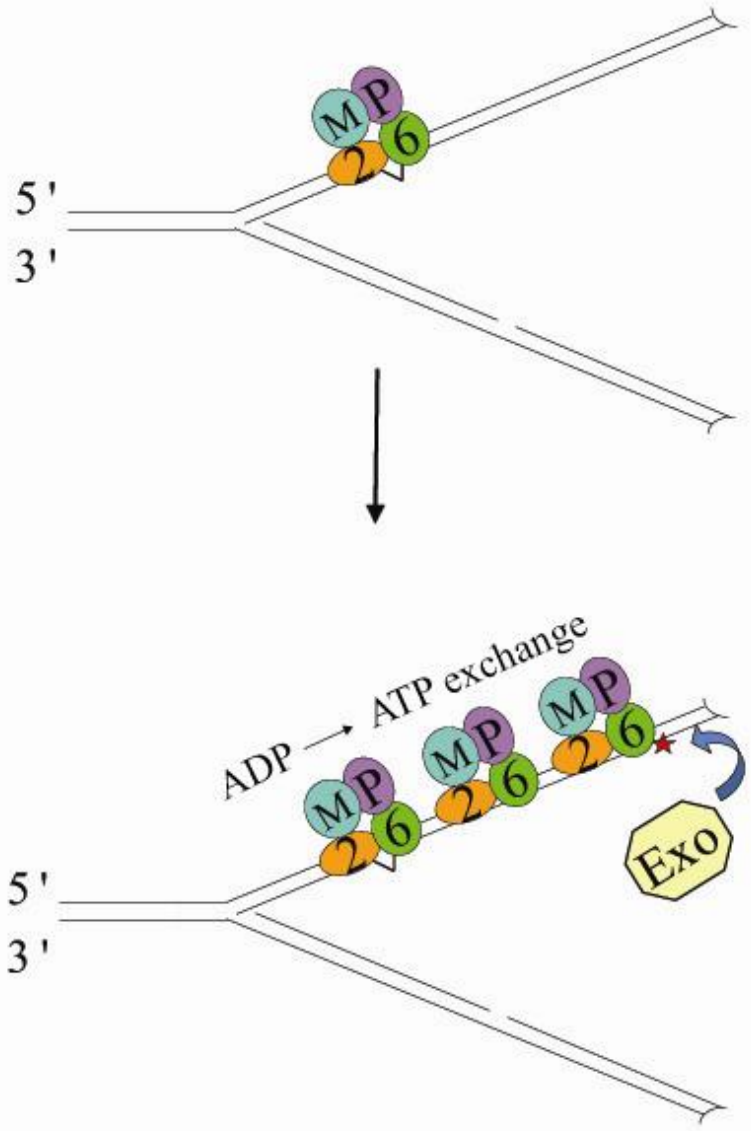
form a ternary complex. In the absence of a mismatch or ATP, MutSα and MutLα do not interact. ATP-dependent dimerization of *E. coli* MutL is also a prerequisite for interactions with MutH and UvrD helicase (Hall *et al.* 1998; Hall and Matson 1999).

What is the significance of the ATP-dependent conformational changes in MLH proteins? The ring-like structural topology of the MLH heterodimer in the presence of ATP suggests that DNA can thread through the central pore of the complex; such a structure could facilitate or regulate the sliding clamp activities of the MSH complex (Figure 1.4). Consistent with this idea, biochemical analysis of MLH proteins has shown that they can bind to DNA; however these proteins do not show an increased affinity for mismatched DNA. DNA binding by MLH proteins appears to be through non-specific electrostatic contacts to the DNA backbone; such binding can be seen on both single-stranded and double-stranded DNA. There is also evidence for cooperative binding and polymerization along the DNA duplex (Hall *et al.* 2001; Gorman *et al.* 2010).

It is unclear what role MLH DNA binding has in bridging mismatch recognition to downstream repair events. This activity, however, may be related to MutLα displaying a latent DNA endonuclease activity that is restricted to DNA near the mismatch site (Kadyrov *et al.* 2006). This activity is essential for MMR and likely acts at the excision step of the process (see Excision section). The endonuclease activity of MutLα resides in the PMS2 subunit and requires both ATP and heavy metal binding. Such an activity was not identified in MutL proteins from bacterial species that contain MutH and use the methylation-dependent strand discrimination mechanism. Intriguingly this observation hints at the possibility that the endonuclease activity of MutLα has replaced the requirement for MutH in higher organisms, though this has yet to be proven.

Figure 1.4 Model for signaling downstream repair.
The Molecular Switch Model proposes that after mismatch recognition, MSH-MLH complexes act as molecular switches triggered by nucleotide exchange. This exchange allows the complexes to enter a sliding clamp diffusion mode in search of the unknown strand discrimination signal (Gradia *et al*. 1999).

Model for signaling downstream repair



Molecular Switch Model
(Fishel et al.)

Downstream steps in MMR

Downstream steps in MMR include strand discrimination, excision, resynthesis, and ligation. The factors that act in these steps do not appear to be MMR specific. They include (1) exonuclease 1 (EXO1), a factor required in multiple DNA repair pathways that acts in excision; (2) proliferating cell nuclear antigen (PCNA), the replication processivity clamp required at multiple stages of MMR; (3) replication factor C (RFC), the replication clamp loader necessary for directing 3' directed excision and proper loading of PCNA to initiate DNA synthesis; (4) replication protein A (RPA), a ssDNA binding protein that plays an important role in stimulating strand excision; (5) high mobility group protein B1 (HMGB1), a non-histone chromatin associated protein likely having redundant functions with RPA; (6) DNA pol δ, a highly accurate polymerase used during resynthesis that fills in the gap created by excision; and (7) DNA ligase I, which seals the nick leftover after resynthesis to form an unbroken strand and complete repair (Kunkel and Erie 2005). Coordination between each of these factors and each step of the MMR pathway is critical for efficient repair; the little that is known about how this is accomplished is indicated below. Mechanisms for strand discrimination, which are still being worked out, will be discussed at the end of this section.

Excision

In reconstituted mammalian MMR involving a nicked mismatch substrate (typically a G-T mismatch), mismatch-dependent excision can be observed in the 5' to 3' or 3' to 5' direction, depending on the position of the strand nick in the DNA substrate relative to the mismatch. MutSα, RPA, and the 5' to 3' double stranded exonuclease EXO1 are sufficient for 5' to 3' excision from a nick to a mismatch; however, MutLα enhances mismatch dependence of this reaction by suppressing excision on homoduplex DNA by EXO1 (Genschel and

Modrich 2003). Recent *in vitro* studies suggest that the addition of PCNA and RFC to a 3'-directed mismatch substrate is sufficient to promote 3' to 5' excision repair and to suppress inappropriate 5' to 3' excision (Zhang *et al.* 2005). These observations raise the question of how EXO1, a 5' to 3' exonuclease, can support both 5' and 3'-directed MMR. The finding that MutLα has a latent endonuclease activity that appears to be restricted to regions surrounding the mismatch provides an elegant solution; the endonuclease cleavages generated by MutLα are thought to provide an entry site for the 5' to 3' EXO1 exonuclease to act in bi-directional MMR.

Resynthesis

During reconstituted mammalian MMR single-strand gaps are created during excision steps; these gaps can be as large as 1 KB in length and must be filled in by DNA polymerase and sealed by DNA ligase I to complete repair (Longley *et al.* 1997; Zhang *et al.* 2005). The gaps created by excision span from a nick site introduced into a plasmid substrate (in lieu of a strand-specific repair signal) to roughly 150 base pairs beyond the mismatch site (Grilley *et al.* 1993). The signaling mechanisms involved in coordinating excision and resynthesis steps are currently being worked out. Many of the MMR factors that act in early steps in MMR (MutSα, MutLα, RPA, PCNA, and RFC) have been implicated in both suppressing and terminating EXO1-mediated excision. The DNA synthesis steps in the reconstituted system are initiated by recruitment of DNA pol δ by PCNA that has been loaded by RFC (Umar *et al.* 1996).

Strand discrimination: the Holy Grail in the MMR field

Unlike many DNA repair events, where a lesion can provide localized information to

direct repair, MMR uses information provided by the DNA replication machinery to remove the mismatch on the newly replicated strand. Perhaps the most interesting question remaining in the mammalian MMR field is how strand discrimination occurs. The nature of the strand discrimination signal is unclear but genetic and biochemical evidence obtained primarily from yeast and bacteria suggest that it involves the identification of nicks on the newly replicated daughter strand (Figure 1.5).

The replication processivity clamp PCNA interacts with many proteins, especially those involved in DNA repair, and this list includes the MutSα, MutSβ, MutLα, and EXO1 MMR factors (Umar *et al.* 1996; Gu *et al.* 1998; Clark *et al.* 2000; Flores-Rozas *et al.* 2000; Bowers *et al.* 2001; Kleczkowska *et al.* 2001; Lau and Kolodner 2003; Dzantiev *et al.* 2004). In one model, PCNA provides a link between MMR and replication to allow mismatches to be identified immediately after formation by concentrating MMR factors at the replication fork. Like the hemi-methylation status of GATC sites in *E. coli*, nicks generated during replication are transient in nature. The identification of these nicks by MMR could thus be used as a strand discrimination signal. PCNA is known to increase mismatch specificity of MutSα to activate EXO1 during excision and is required to initiate resynthesis by DNA polymerase (Flores-Rozas *et al.* 2000; Genschel and Modrich 2003). Furthermore, PCNA is always loaded onto junctions of single and double stranded DNA by RFC in a specific orientation. Thus PCNA by virtue of its orientation upon loading could relay strand information through interactions with MutSα or MutLα (Bowman *et al.* 2004). After MutSα or MutLα is oriented onto DNA, they could search out nearby strand discontinuities created during replication (Pluciennik *et al.* 2010). There is also the possibility that a PCNA-independent mechanism orients MutSα or MutLα.

Depending on the orientation of the strand discontinuity to the mismatch, excision can

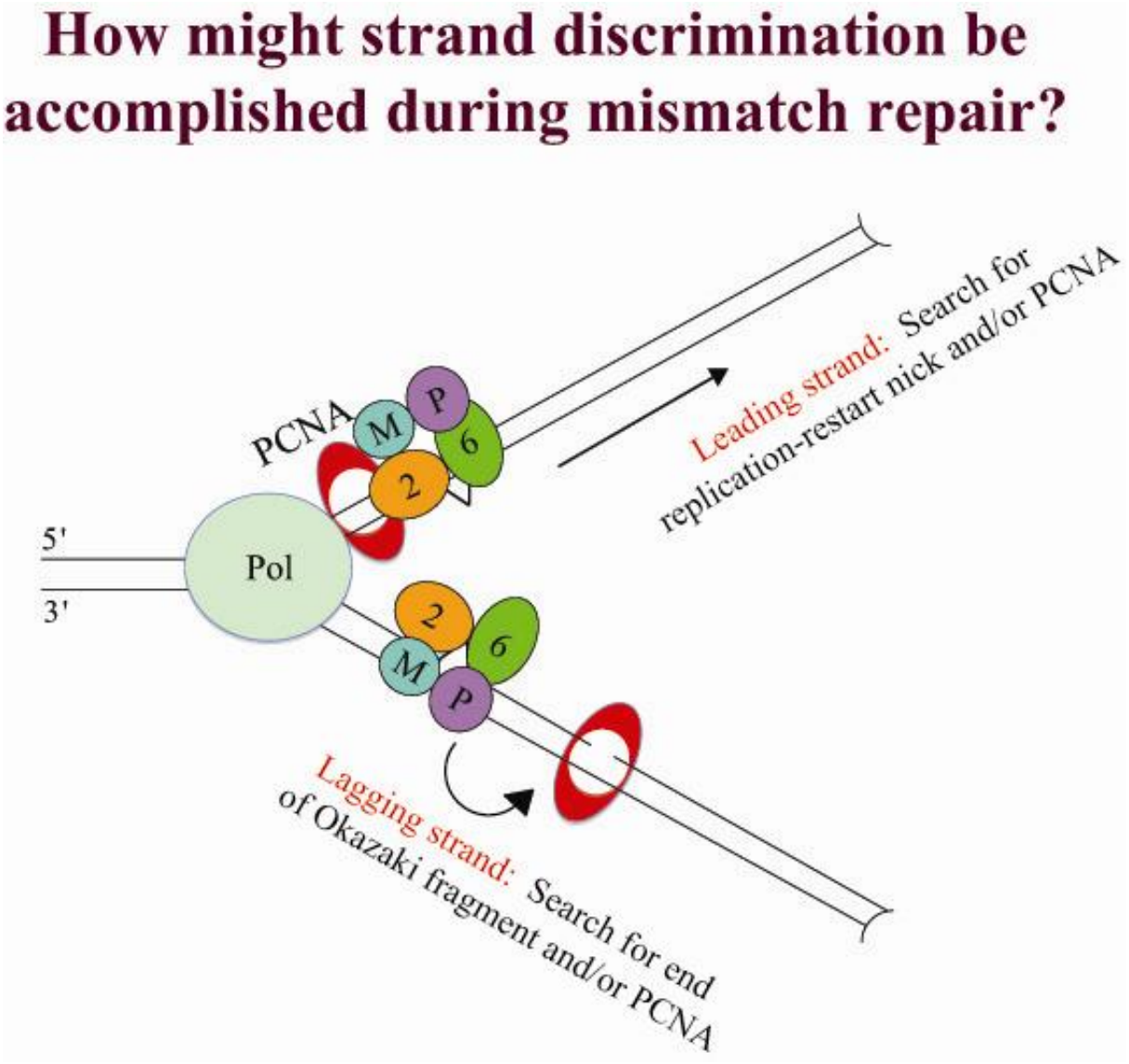


Figure 1.5 How might strand discrimination be accomplished during mismatch repair? Shown is a model of a replication fork with mismatches indicated by triangles. After mismatches are recognized, repair is directed to the daughter strand. Orientation of MutSα (MSH2-MSH6) or MutLα (MLH1-PMS2) to initiate excision of the newly replicated strand may be facilitated by interactions with PCNA. The nature of the strand discrimination signal is unclear but is likely a nick created during replication. The lagging strand (bottom strand) contains unprocessed Okazaki fragments that provide nicks. The leading strand (top strand) would only generate nicks if the replication fork was restarted. Alternatively it is possible that the DNA end at the elongating replication fork may serve to discriminate strands. MutLα could also be directed to cleave the correct strand to promote excision in each case.

be directly started by the 5' to 3' exonucleases activity of EXO1 or, to allow 3'-directed repair by EXO1, would require a new strand break to be created on the opposite side of the mismatch through MutLα endonucleolytic cleavage. This model is supported by *in vitro* data showing that MutLα is dispensable on a substrate with a nick 5' to a mismatch, but is absolutely required for a substrate with a nick 3' to a mismatch (Constantin *et al.* 2005). In order for the MMR system to utilize this mechanism, it would require relatively frequent strand discontinuities to be created during replication. Such discontinuities occur during lagging strand synthesis where the average Okazaki fragment size is a few hundred base pairs (Waga and Stillman 1998). The fact that MMR is observed to be more efficient on the lagging strand may reflect the high frequency of nicks that are accessible to MMR factors on this strand. There is little evidence for the presence of nicks on the leading strand at a density (every few KB) that would be utilized by MMR. However, several studies have suggested that leading strand synthesis may not be as continuous as had been thought (Wang and Chen 1994; Heller and Marians 2006). One possibility is that like lagging strand synthesis, replication is regularly restarted on the leading strand, thus creating strand discontinuities that can serve as a strand discrimination signal. There is also the possibility that other undiscovered factors or signals play a role in this process.

Another point of contention in the field is how identification of the newly replicated strand is coordinated with mismatch recognition and downstream signaling. Recent evidence in bacterial and mammalian MMR systems suggests that MSH proteins form a sliding clamp with MLH proteins (Figure 1.4). Experiments in *E. coli* suggest that after mismatch recognition, MutS and MutL track along the helical contour to activate MutH to cleave the nearest hemi-methylated GATC site (Pluciennik and Modrich 2007). These studies support a model in mammalian cells proposed by Rick Fishel and colleagues in which, following

mismatch recognition, MSH-MLH complexes acts as molecular switches that are triggered by nucleotide exchange to enter a sliding clamp diffusion mode (Gradia *et al.* 1997). In this model MSH-MLH complexes then locate presently unknown strand discrimination signals and guide excision steps that result in mismatch removal. An implication of this model is that multiple binding and release events can take place at the mismatch that may lead to polymer formation and communication between the two sites. The fact that MutL proteins can form polymers on DNA provides support for this idea (Hall *et al.* 2001). Excision needs to take place in the correct direction for competent repair to occur. Therefore, communication between the mismatch and the strand discrimination signal is likely to be important because of the bi-directional nature of MMR.

Dissertation outline

This dissertation concentrates on several properties of MMR proteins that contribute to their roles in MMR and meiotic recombination in *Saccharomyces cerevisiae*.

Chapter 2 describes recently published work on the diffusion properties of purified yeast MMR proteins along DNA. The ability of MLH1-PMS1 to traverse barriers, such as nucleosomes, is presented. This work demonstrates for the first time that a passively diffusing protein, MLH1-PMS1, can traverse stationary nucleosomes.

Chapter 3 analyzes the importance of unstructured linker domains of MLH1 and PMS1 in MMR. This work is extended from findings in Chapter 2 that begin to characterize the mechanistic details of MLH1-PMS1 diffusion. My work demonstrates that intact linker arms in MLH1-PMS1 are critical for DNA binding and that shortening these linker arms disrupts MMR. This chapter also includes future directions for this project that focuses on determining the diffusion properties of mutant proteins presented.

Chapter 4 details both published and unpublished work aimed at characterizing the role for a putative endonuclease domain of MLH1-MLH3 in MMR and meiotic recombination. Previous genetic analysis suggested that mutations in the putative endonuclease domain of MLH1-MLH3 resulted in a defect in MMR and meiotic recombination and led to the hypothesis that MLH1-MLH3 is a Holiday Junction Resolvase. My work discovered that MLH1-MLH3 was difficult to purify and characterize using a variety of purification techniques. Partially purified complexes did not display an endonuclease activity. Further studies will be necessary to confirm these negative results.

The Appendix describes supplementary data from Chapters 2 and 3. For Chapter 2, this includes supplementary discussion regarding the effects of salt on facilitated diffusion

and supplementary figures that are controls to confirm the diffusion characteristics of MLH1-PMS1. For Chapter 3, the different complexes of engineered N-terminal swap constructs that were tested in MMR assays are illustrated.

References

Alani, E., J. Y. Lee, et al. (2003). "Crystal structure and biochemical analysis of the MutS.ADP.beryllium fluoride complex suggests a conserved mechanism for ATP interactions in mismatch repair." J Biol Chem **278**(18): 16088-94.

Alani, E., T. Sokolsky, et al. (1997). "Genetic and biochemical analysis of Msh2p-Msh6p: role of ATP hydrolysis and Msh2p-Msh6p subunit interactions in mismatch base pair recognition." Mol Cell Biol **17**(5): 2436-47.

Arana, M. E., S. F. Holmes, et al. (2010). "Functional residues on the surface of the N-terminal domain of yeast Pms1." DNA Repair (Amst) **9**(4): 448-57.

Au, K. G., K. Welsh, et al. (1992). "Initiation of methyl-directed mismatch repair." J Biol Chem **267**(17): 12142-8.

Ban, C., M. Junop, et al. (1999). "Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair." Cell **97**(1): 85-97.

Ban, C. and W. Yang (1998). "Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis." Cell **95**(4): 541-52.

Bellacosa, A. (2001). "Functional interactions and signaling properties of mammalian DNA mismatch repair proteins." Cell Death Differ **8**(11): 1076-92.

Bowers, J., P. T. Tran, et al. (2001). "MSH-MLH complexes formed at a DNA mismatch are disrupted by the PCNA sliding clamp." J Mol Biol **306**(5): 957-68.

Bowman, G. D., M. O'Donnell, et al. (2004). "Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex." Nature **429**(6993): 724-30.

Brooks, J. E., R. M. Blumenthal, et al. (1983). "The isolation and characterization of the Escherichia coli DNA adenine methylase (dam) gene." Nucleic Acids Res **11**(3): 837-51.

Buermeyer, A. B., S. M. Deschenes, et al. (1999). "Mammalian DNA mismatch repair." Annu Rev Genet **33**: 533-64.

Burdett, V., C. Baitinger, et al. (2001). "In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair." Proc Natl Acad Sci U S A **98**(12): 6765-70.

Chen, S., S. H. Bigner, et al. (2001). "High rate of CAD gene amplification in human cells deficient in MLH1 or MSH6." Proc Natl Acad Sci U S A **98**(24): 13802-7.

Chi, N. W. and R. D. Kolodner (1994). "Purification and characterization of MSH1, a yeast mitochondrial protein that binds to DNA mismatches." J Biol Chem **269**(47): 29984-92.

Clark, A. B., F. Valle, et al. (2000). "Functional interaction of proliferating cell nuclear antigen with MSH2-MSH6 and MSH2-MSH3 complexes." J Biol Chem **275**(47): 36498-501.

Constantin, N., L. Dzantiev, et al. (2005). "Human mismatch repair: reconstitution of a nick-directed bidirectional reaction." J Biol Chem **280**(48): 39752-61.

Drotschmann, K., W. Yang, et al. (2001). "Asymmetric recognition of DNA local distortion. Structure-based functional studies of eukaryotic Msh2-Msh6." J Biol Chem **276**(49): 46225-9.

Dutta, R. and M. Inouye (2000). "GHKL, an emergent ATPase/kinase superfamily." Trends Biochem Sci **25**(1): 24-8.

Dzantiev, L., N. Constantin, et al. (2004). "A defined human system that supports bidirectional mismatch-provoked excision." Mol Cell **15**(1): 31-41.

Flores-Rozas, H., D. Clark, et al. (2000). "Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mispair recognition complex." Nat Genet **26**(3): 375-8.

Genschel, J., L. R. Bazemore, et al. (2002). "Human exonuclease I is required for 5' and 3' mismatch repair." J Biol Chem **277**(15): 13302-11.

Genschel, J. and P. Modrich (2003). "Mechanism of 5'-directed excision in human mismatch repair." Mol Cell **12**(5): 1077-86.

Glickman, B. W. and M. Radman (1980). "Escherichia coli mutator mutants deficient in methylation-instructed DNA mismatch correction." Proc Natl Acad Sci U S A **77**(2): 1063-7.

Gorman, J., A. Chowdhury, et al. (2007). "Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6." Mol Cell **28**(3): 359-70.

Gorman, J., A. J. Plys, et al. (2010). "Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice." Nat Struct Mol Biol **17**(8): 932-8.

Gradia, S., S. Acharya, et al. (1997). "The human mismatch recognition complex hMSH2-hMSH6 functions as a novel molecular switch." Cell **91**(7): 995-1005.

Gradia, S., D. Subramanian, et al. (1999). "hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA." Mol Cell **3**(2): 255-61.

Grilley, M., J. Griffith, et al. (1993). "Bidirectional excision in methyl-directed mismatch repair." J Biol Chem **268**(16): 11830-7.

Gu, L., Y. Hong, et al. (1998). "ATP-dependent interaction of human mismatch repair proteins and dual role of PCNA in mismatch repair." Nucleic Acids Res **26**(5): 1173-8.

Guarne, A., M. S. Junop, et al. (2001). "Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase." Embo J **20**(19): 5521-31.

Guarne, A., S. Ramon-Maiques, et al. (2004). "Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair." Embo J **23**(21): 4134-45.

Habraken, Y., P. Sung, et al. (1996). "Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3." Curr Biol **6**(9): 1185-7.

Habraken, Y., P. Sung, et al. (1998). "ATP-dependent assembly of a ternary complex consisting of a DNA mismatch and the yeast MSH2-MSH6 and MLH1-PMS1 protein complexes." J Biol Chem **273**(16): 9837-41.

Hall, M. C., J. R. Jordan, et al. (1998). "Evidence for a physical interaction between the Escherichia coli methyl-directed mismatch repair proteins MutL and UvrD." Embo J **17**(5): 1535-41.

Hall, M. C. and S. W. Matson (1999). "The Escherichia coli MutL protein physically interacts with MutH and stimulates the MutH-associated endonuclease activity." J Biol Chem **274**(3): 1306-12.

Hall, M. C., P. V. Shcherbakova, et al. (2002). "Differential ATP binding and intrinsic ATP hydrolysis by amino-terminal domains of the yeast Mlh1 and Pms1 proteins." J Biol Chem **277**(5): 3673-9.

Hall, M. C., H. Wang, et al. (2001). "High affinity cooperative DNA binding by the yeast Mlh1-Pms1 heterodimer." J Mol Biol **312**(4): 637-47.

Harfe, B. D. and S. Jinks-Robertson (2000). "DNA mismatch repair and genetic instability." Annu Rev Genet **34**: 359-399.

Harfe, B. D., B. K. Minesinger, et al. (2000). "Discrete in vivo roles for the MutL homologs Mlh2p and Mlh3p in the removal of frameshift intermediates in budding yeast." Curr Biol **10**(3): 145-8.

Hattman, S., J. E. Brooks, et al. (1978). "Sequence specificity of the P1 modification methylase (M.Eco P1) and the DNA methylase (M.Eco dam) controlled by the Escherichia coli dam gene." J Mol Biol **126**(3): 367-80.

Heller, R. C. and K. J. Marians (2006). "Replication fork reactivation downstream of a blocked nascent leading strand." Nature **439**(7076): 557-62.

Hollingsworth, N. M., L. Ponte, et al. (1995). "MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in Saccharomyces cerevisiae but not mismatch repair." Genes Dev **9**(14): 1728-39.

Iaccarino, I., F. Palombo, et al. (1996). "MSH6, a Saccharomyces cerevisiae protein that binds to mismatches as a heterodimer with MSH2." Curr Biol **6**(4): 484-6.

Irving, J. A. and A. G. Hall (2001). "Mismatch repair defects as a cause of resistance to cytotoxic drugs." Expert Rev Anticancer Ther **1**(1): 149-58.

Iyer, R. R., A. Pluciennik, et al. (2006). "DNA mismatch repair: functions and mechanisms." Chem Rev **106**(2): 302-23.

Jiricny, J. (2006). "The multifaceted mismatch-repair system." Nat Rev Mol Cell Biol **7**(5): 335-46.

Kadyrov, F. A., L. Dzantiev, et al. (2006). "Endonucleolytic function of MutLalpha in human mismatch repair." Cell **126**(2): 297-308.

Kleczkowska, H. E., G. Marra, et al. (2001). "hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci." Genes Dev **15**(6): 724-36.

Kolodner, R. D. and G. T. Marsischky (1999). "Eukaryotic DNA mismatch repair." Curr Opin Genet Dev **9**(1): 89-96.

Kosinski, J., I. Steindorf, et al. (2005). "Analysis of the quaternary structure of the MutL C-terminal domain." J Mol Biol **351**(4): 895-909.

Kunkel, T. A. and D. A. Erie (2005). "DNA mismatch repair." Annu Rev Biochem **74**: 681-710.

Lahue, R. S., K. G. Au, et al. (1989). "DNA mismatch correction in a defined system." Science **245**(4914): 160-4.

Lamers, M. H., A. Perrakis, et al. (2000). "The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch." Nature **407**(6805): 711-7.

Lamers, M. H., H. H. Winterwerp, et al. (2003). "The alternating ATPase domains of MutS control DNA mismatch repair." Embo J **22**(3): 746-56.

Lau, P. J. and R. D. Kolodner (2003). "Transfer of the MSH2.MSH6 complex from proliferating cell nuclear antigen to mispaired bases in DNA." J Biol Chem **278**(1): 14-7.

Li, G. M. (2008). "Mechanisms and functions of DNA mismatch repair." Cell Res **18**(1): 85-98.

Liu, B., R. Parsons, et al. (1996). "Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients." Nat Med **2**(2): 169-74.

Longley, M. J., A. J. Pierce, et al. (1997). "DNA polymerase delta is required for human mismatch repair in vitro." J Biol Chem **272**(16): 10917-21.

Lynch, H. T., P. M. Lynch, et al. (2009). "Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications." Clin Genet **76**(1): 1-18.

Marti, T. M., C. Kunz, et al. (2002). "DNA mismatch repair and mutation avoidance pathways." J Cell Physiol **191**(1): 28-41.

Martik, D., C. Baitinger, et al. (2004). "Differential specificities and simultaneous occupancy of human MutSalpha nucleotide binding sites." J Biol Chem **279**(27): 28402-10.

Martin, A. and M. D. Scharff (2002). "AID and mismatch repair in antibody diversification." Nat Rev Immunol **2**(8): 605-14.

Matson, S. W. and A. B. Robertson (2006). "The UvrD helicase and its modulation by the mismatch repair protein MutL." Nucleic Acids Res **34**(15): 4089-97.

Mazur, D. J., M. L. Mendillo, et al. (2006). "Inhibition of Msh6 ATPase activity by mispaired DNA induces a Msh2(ATP)-Msh6(ATP) state capable of hydrolysis-independent movement along DNA." Mol Cell **22**(1): 39-49.

Messer, W. and M. Noyer-Weidner (1988). "Timing and targeting: the biological functions of Dam methylation in E. coli." Cell **54**(6): 735-7.

Modrich, P. and R. Lahue (1996). "Mismatch repair in replication fidelity, genetic recombination, and cancer biology." Annu Rev Biochem **65**: 101-33.

Natrajan, G., M. H. Lamers, et al. (2003). "Structures of Escherichia coli DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates." Nucleic Acids Res **31**(16): 4814-21.

Obmolova, G., C. Ban, et al. (2000). "Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA." Nature **407**(6805): 703-10.

Palombo, F., I. Iaccarino, et al. (1996). "hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA." Curr Biol **6**(9): 1181-4.

Pluciennik, A., L. Dzantiev, et al. (2010). "PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair." Proc Natl Acad Sci U S A **107**(37): 16066-71.

Pluciennik, A. and P. Modrich (2007). "Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair." Proc Natl Acad Sci U S A **104**(31): 12709-13.

Raschle, M., G. Marra, et al. (1999). "Identification of hMutLbeta, a heterodimer of hMLH1 and hPMS1." J Biol Chem **274**(45): 32368-75.

Runyon, G. T., D. G. Bear, et al. (1990). "Escherichia coli helicase II (UvrD) protein initiates DNA unwinding at nicks and blunt ends." Proc Natl Acad Sci U S A **87**(16): 6383-7.

Sacho, E. J., F. A. Kadyrov, et al. (2008). "Direct visualization of asymmetric adenine-nucleotide-induced conformational changes in MutL alpha." Mol Cell **29**(1): 112-21.

Studamire, B., T. Quach, et al. (1998). "Saccharomyces cerevisiae Msh2p and Msh6p ATPase activities are both required during mismatch repair." Mol Cell Biol **18**(12): 7590-601.

Tessmer, I., Y. Yang, et al. (2008). "Mechanism of MutS searching for DNA mismatches and signaling repair." J Biol Chem **283**(52): 36646-54.

Tran, H. T., D. A. Gordenin, et al. (1999). "The 3'-->5' exonucleases of DNA polymerases delta and epsilon and the 5'-->3' exonuclease Exo1 have major roles in postreplication mutation avoidance in Saccharomyces cerevisiae." Mol Cell Biol **19**(3): 2000-7.

Tran, P. T. and R. M. Liskay (2000). "Functional studies on the candidate ATPase domains of Saccharomyces cerevisiae MutLalpha." Mol Cell Biol **20**(17): 6390-8.

Umar, A., A. B. Buermeyer, et al. (1996). "Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis." Cell **87**(1): 65-73.

Viswanathan, M., V. Burdett, et al. (2001). "Redundant exonuclease involvement in Escherichia coli methyl-directed mismatch repair." J Biol Chem **276**(33): 31053-8.

Waga, S. and B. Stillman (1998). "The DNA replication fork in eukaryotic cells." Annu Rev Biochem **67**: 721-51.

Wang, H., Y. Yang, et al. (2003). "DNA bending and unbending by MutS govern mismatch recognition and specificity." Proc Natl Acad Sci U S A **100**(25): 14822-7.

Wang, T. C. and S. H. Chen (1994). "Okazaki DNA fragments contain equal amounts of lagging-strand and leading-strand sequences." Biochem Biophys Res Commun **198**(3): 844-9.

Wang, T. F., N. Kleckner, et al. (1999). "Functional specificity of MutL homologs in yeast: evidence for three Mlh1-based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction." Proc Natl Acad Sci U S A **96**(24): 13914-9.

Warren, J. J., T. J. Pohlhaus, et al. (2007). "Structure of the human MutSalpha DNA lesion recognition complex." Mol Cell **26**(4): 579-92.

Welsh, K. M., A. L. Lu, et al. (1987). "Isolation and characterization of the Escherichia coli mutH gene product." J Biol Chem **262**(32): 15624-9.

Zhang, Y., F. Yuan, et al. (2005). "Reconstitution of 5'-directed human mismatch repair in a purified system." Cell **122**(5): 693-705.

Chapter 2

Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice

Aaron J. Plys[1,5], Jason Gorman[2,5], Mari-Liis Visnapuu[3], Eric Alani[1] and Eric C. Greene[3,4]

[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York, USA. [2]Department of Biological Sciences, Columbia University, New York, New York, USA. [3]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York, USA. [4]Howard Hughes Medical Institute, Columbia University, New York, New York, USA. [5]These authors contributed equally to this work.

This chapter was originally published in the August 2010 issue of Nature Structural & Molecular Biology:

Contributions:

A.J.P. designed and engineered all constructs, expressed and purified the proteins, did all ensemble-level characterization and conducted the immunoprecipitation experiments; J.G. designed the TIRFM experiments and collected and analyzed the single-molecule data; M.-L.V. made and characterized all chromatin substrates; A.J.P., J.G., M.-L.V., E.A. and E.C.G. discussed the data and cowrote the paper.

Gorman, Jason, Aaron J. Plys, Mari-Liis Visnapuu, Eric Alani, and Eric Greene. 2010: Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. Nature Structural & Molecular Biology. 2010. Volume 17, Issue 8, Pages 932-938.

Abstract

DNA-binding proteins survey genomes for targets using facilitated diffusion, which typically includes a one-dimensional (1D) scanning component for sampling local regions. Eukaryotic proteins must accomplish this task while navigating through chromatin. Yet it is unknown whether nucleosomes disrupt 1D scanning or eukaryotic DNA-binding factors can circumnavigate nucleosomes without falling off DNA. Here we use single-molecule microscopy in conjunction with nanofabricated curtains of DNA to show that the postreplicative mismatch repair protein complex Mlh1–Pms1 diffuses in 1D along DNA via a hopping/stepping mechanism and readily bypasses nucleosomes. This is the first experimental demonstration that a passively diffusing protein can traverse stationary obstacles. In contrast, Msh2–Msh6, a mismatch repair protein complex that slides while maintaining continuous contact with DNA, experiences a boundary upon encountering nucleosomes. These differences reveal important mechanistic constraints affecting intranuclear trafficking of DNA-binding proteins.

Introduction

Virtually all DNA-binding proteins must use some form of facilitated diffusion (for example, hopping, jumping, sliding and/or intersegmental transfer) to scan the genome and locate targets (von Hippel and Berg 1989; Elf *et al.* 2007; Hager *et al.* 2009; Li *et al.* 2009). The advent of single-molecule imaging has led to a resurgence of interest in facilitated diffusion, and an emerging consensus agrees that many proteins can scan DNA via one-dimensional (1D) diffusion, where the proteins undergo a random walk while moving laterally along the helix (Blainey *et al.* 2006; Gorski *et al.* 2006; Elf *et al.* 2007; Gorman and Greene 2008; Hager *et al.* 2009; Li *et al.* 2009). However, all of these studies have been limited to naked DNA substrates, which do not resemble the crowded environments that would be encountered *in vivo*, leaving the role of 1D diffusion in question under physiologically relevant settings (Gorski *et al.* 2006; Hager *et al.* 2009; Li *et al.* 2009). In eukaryotes, these processes must occur within the context of chromatin, which has the potential to hinder protein mobility (Kampmann 2005; Gorski *et al.* 2006; Hager *et al.* 2009; Li *et al.* 2009). Motor proteins, such as RNA polymerase and other DNA translocases, solve this problem by using the chemomechanical energy derived from nucleotide hydrolysis to push their way through nucleosome obstacles (Studitsky *et al.* 1995; Hodges *et al.* 2009). However, most DNA-binding proteins, such as transcription factors or DNA-repair proteins, cannot mechanically disrupt nucleosomes; therefore, other mechanisms must come into play if these proteins are to scan chromatin. Whether proteins can circumnavigate nucleosomes without dissociating from DNA remains an unresolved issue with direct bearing on how all eukaryotic DNA-binding proteins are trafficked throughout the nucleus (Kampmann 2005; Gorski *et al.*

2006; Hager *et al.* 2009; Li *et al.* 2009). This problem led us to ask whether eukaryotic proteins that diffuse in 1D along DNA could circumnavigate individual nucleosomes and travel along nucleosomal arrays, and if so, what mechanistic principles affect mobility along chromatin.

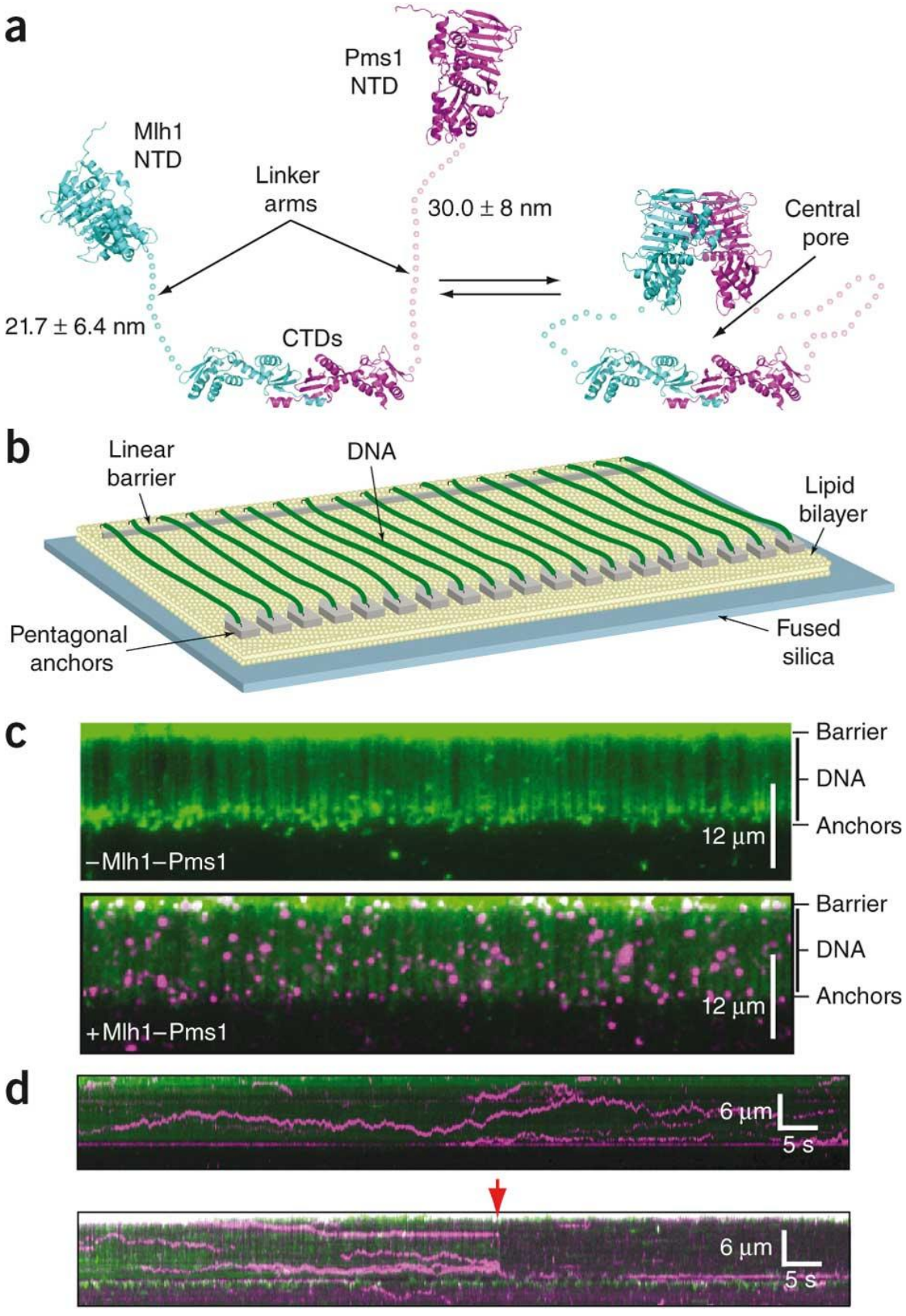
We chose the *Saccharomyces cerevisiae* post-replicative mismatch repair (MMR) protein complexes Msh2–Msh6 and Mlh1–Pms1 as model systems for studying the physical basis of facilitated diffusion. MMR is a ubiquitous repair pathway that corrects errors (mismatches and small insertion/deletion loops) left behind by the replication machinery (Kunkel and Erie 2005; Jiricny 2006; Modrich 2006). Defects in MMR lead to elevated mutation rates, are linked to hereditary nonpolyposis colon cancer (HNPCC) and are associated with many sporadic tumors (Modrich 2006). Msh2–Msh6 and Mlh1–Pms1 are DNA-binding proteins required for MMR. During MMR, Msh2–Msh6 must locate lesions and also helps identify nearby signals differentiating parental and nascent DNA strands, whereas Mlh1–Pms1 must locate lesion-bound Msh2–Msh6 and then coordinates downstream steps in the reaction. Although Msh2–Msh6 and Mlh1–Pms1 are both ATPases, neither uses ATP for generating chemomechanical force; rather, nucleotide binding and hydrolysis are thought to serve as signaling mechanisms for coordinating the various stages of repair by regulating protein-protein interactions in the case of Mlh1–Pms1 or protein-DNA interactions with Msh2–Msh6 (Kunkel and Erie 2005; Jiricny 2006). These or closely related protein complexes are also involved in mitotic and meiotic recombination, triplet-repeat expansion, class-switch recombination, somatic hypermutation and DNA-damage signaling checkpoints (Jiricny 2006). All known functions of Msh2–Msh6 and Mlh1–Pms1 require targeting to specific structures within the genome, and the later stages of the MMR reaction involved in

strand discrimination are also thought to involve 1D movement along DNA (Kunkel and Erie 2005; Jiricny 2006; Modrich 2006), making these protein complexes good candidates as model systems for single-molecule studies of facilitated diffusion.

Results

Experimental approach for visualizing protein-DNA interactions. Using total internal reflection fluorescent microscopy (TIRFM) we have previously shown that Msh2–Msh6 moves on DNA via a sliding mechanism consistent with a model where it tracks the phosphate backbone (Gorman *et al.* 2007). To determine whether Mlh1–Pms1 also moves on DNA, we engineered the proteins (Figure 2.1) with epitope tags (flag and/or hemagglutinin) and labeled them with antibody-coupled quantum dots (QDs). Gel shift and nitrocellulose filter-binding assays confirmed labeling specificity and showed that labeling did not disrupt DNA binding activity (Figure 2.2A-C). For TIRFM, we used microfluidic devices with hybrid surfaces comprised of fluid lipid bilayers and nanofabricated metallic barrier patterns made by electron-beam lithography (Gorman *et al.* 2010). The DNA substrates (λ-DNA, 48,502 base pairs) were anchored by one end to the bilayer through a biotin-streptavidin linkage, and hydrodynamic force was then used to push the DNA and align it along the leading edges of nanofabricated barriers to lipid diffusion (Figure 2.1B). The second end of the DNA was then anchored to antibody-coated pentagons positioned downstream from the linear barriers (Figure 2.1B). This strategy yields 'double-tethered' curtains of DNA in which the individual DNA molecules are suspended above a lipid bilayer and are anchored by both ends such that they can be viewed across their entire contour length by TIRFM in the absence of a perturbing

Figure 2.1. Nanofabricated racks of DNA for visualizing 1D diffusion of Mlh1–Pms1. **(A)** Illustration of some predicted structures for Mlh1–Pms1 based on structural and biochemical data (Guarne *et al.* 2001; Guarne *et al.* 2004; Kosinski *et al.* 2005; Sacho *et al.* 2008). The NTDs, CTDs, central pore and linker arms are indicated. **(B)** Diagram of the nanofabricated rack device used for making the double-tethered DNA curtains (Gorman *et al.* 2010). The rack consists of linear barriers to lipid diffusion, which align the lipid-tethered DNA molecules, followed by an array of antibody-coated pentagons that provide immobile anchor points for the second end of the DNA. Pattern elements are ~20 nm tall, and the bilayer is ~5 nm thick. **(C)** Top and bottom, YOYO1-stained λ-DNA curtains (green; 48,502 base pairs) with and without QD-tagged Mlh1–Pms1 (magenta), respectively. DNA-bound proteins were not detected in reactions using incorrect antibody-epitope pairs (not shown), and QDs alone did not bind DNA (not shown). **(D)** Kymograms illustrating the motion of Mlh1–Pms1. The lower panel shows photocleavage (arrowhead) of a DNA during data collection. The proteins disappear from view when the DNA breaks, showing that the proteins and DNA are not adsorbed to the bilayer.


a
Pms1
NTD
Mlh1
NTD
Linker
arms
30.0 ± 8 nm
Central
pore
21.7 ± 6.4 nm
CTDs
b
Linear
barrier
DNA
Lipid
bilayer
Pentagonal
anchors
Fused
silica
c
Barrier
DNA
Anchors
12 μm
–Mlh1–Pms1
Barrier
DNA
Anchors
12 μm
+Mlh1–Pms1
d
6 μm
5 s
6 μm
5 s

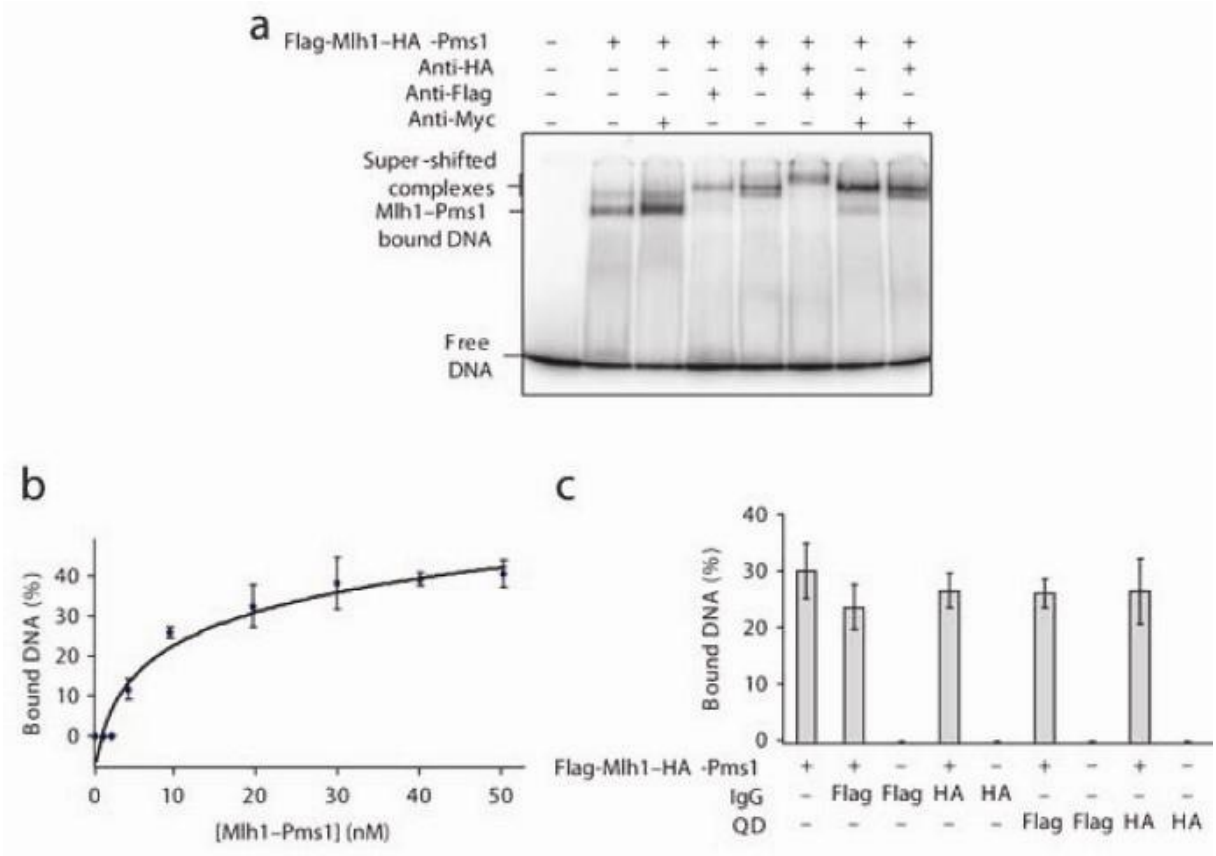


Figure 2.2. DNA-binding activity of fluorescently tagged Mlh1–Pms1.
(A) Gel mobility shifts were performed as described in Materials and Methods. Reactions contained 120 nM Mlh1–Pms1 and 60 nM 5′-^{32}P-labeled 40-mer dsDNA. Antibodies were preincubated with Mlh1–Pms1, as indicated. **(B)** Nitrocellulose filter binding assays were performed with 100 pM 5′-^{32}P-labeled 3-kb linear plasmid at the indicated concentrations of Mlh1–Pms1, and the percentage of bound DNA was determined by dividing the background subtracted counts measured for each filter by the total amount of DNA in the reactions. **(C)** The effects of antibodies and antibody-labeled QDs on the binding activity of Mlh1–Pms1 (20 nM) as determined by the nitrocellulose filter-binding assay. All data points are reported as mean ± s.d.

hydrodynamic flow (Figure 2.1B,C, Appendix, Supplementary Figure 2.7).

Mlh1–Pms1 diffuses on DNA by a stepping or hopping mechanism. When imaged by TIRFM, Mlh1–Pms1 colocalized with DNA (Figure 2.1C), and ≥95% of the DNA-bound proteins moved rapidly back and forth along the DNA molecules (Figure 2.1D, Appendix, Supplementary Table 2.1). Two-color labeling experiments revealed that most (98.4%) of the complexes were single heterodimers under the conditions used for these experiments. Mlh1–Pms1 often remained bound to the DNA for several minutes without dissociating (Figure 2.1), consistent with bulk biochemical studies (Hall *et al.* 2001). Analysis of the motion revealed linear mean-squared displacement (MSD) plots, as expected for 1D diffusion (Blainey *et al.* 2006; Gorman *et al.* 2007) (Figure 2.3A), yielding a mean diffusion coefficient of $D_{1D} = 0.143 \pm 0.29$ μm^2/sec ($N = 25$) at 150 mM NaCl, 1 mM ATP and 1 mM $MgCl_2$ (Figure 2.3B). We observed 1D diffusion with or without ADP, ATP, and ATPγS (Figure 2.3B), and the differences in the diffusion coefficients measured in the presence and absence of ADP or ATP were statistically insignificant (Student's *t*-test, $P \geq 0.01$). These results indicate that nucleotide binding and hydrolysis were unnecessary for movement, consistent with the notion that nucleotide binding is primarily involved in promoting protein-protein interactions or structural rearrangements, with little impact on DNA-binding (Kunkel and Erie 2005; Jiricny 2006; Sacho *et al.* 2008).

Mlh1–Pms1 diffusion coefficients were an order of magnitude greater (Student's *t*-test, $P < 0.0001$) than those of Msh2–Msh6 under physiological salt concentrations (0.143 ± 0.29 μm^2/sec vs. 0.009 ± 0.011 μm^2/sec at 150 mM NaCl; Figure 2.3B) (Gorman *et al.* 2007), suggesting that the two complexes might move via different mechanisms. Potential

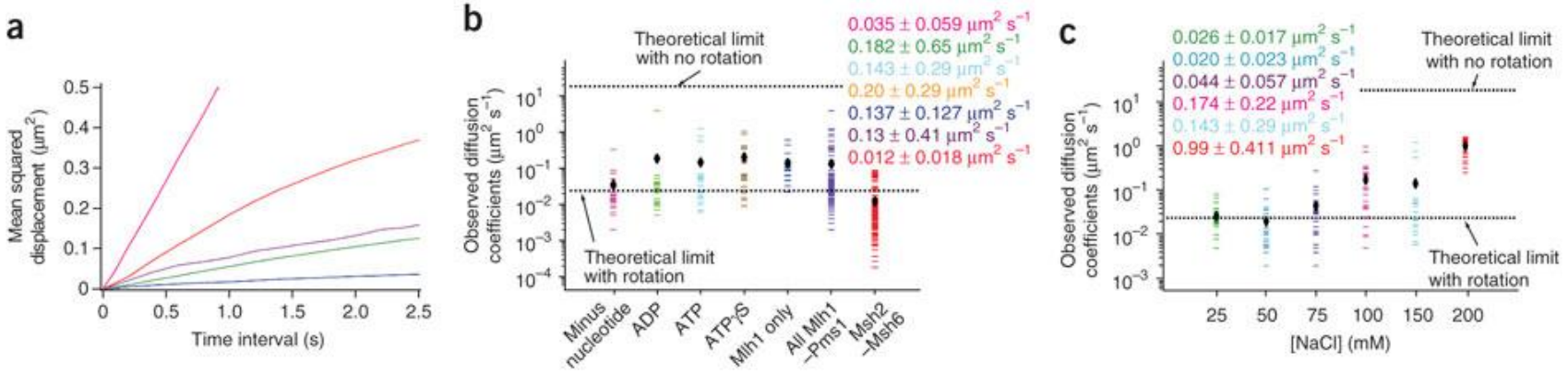


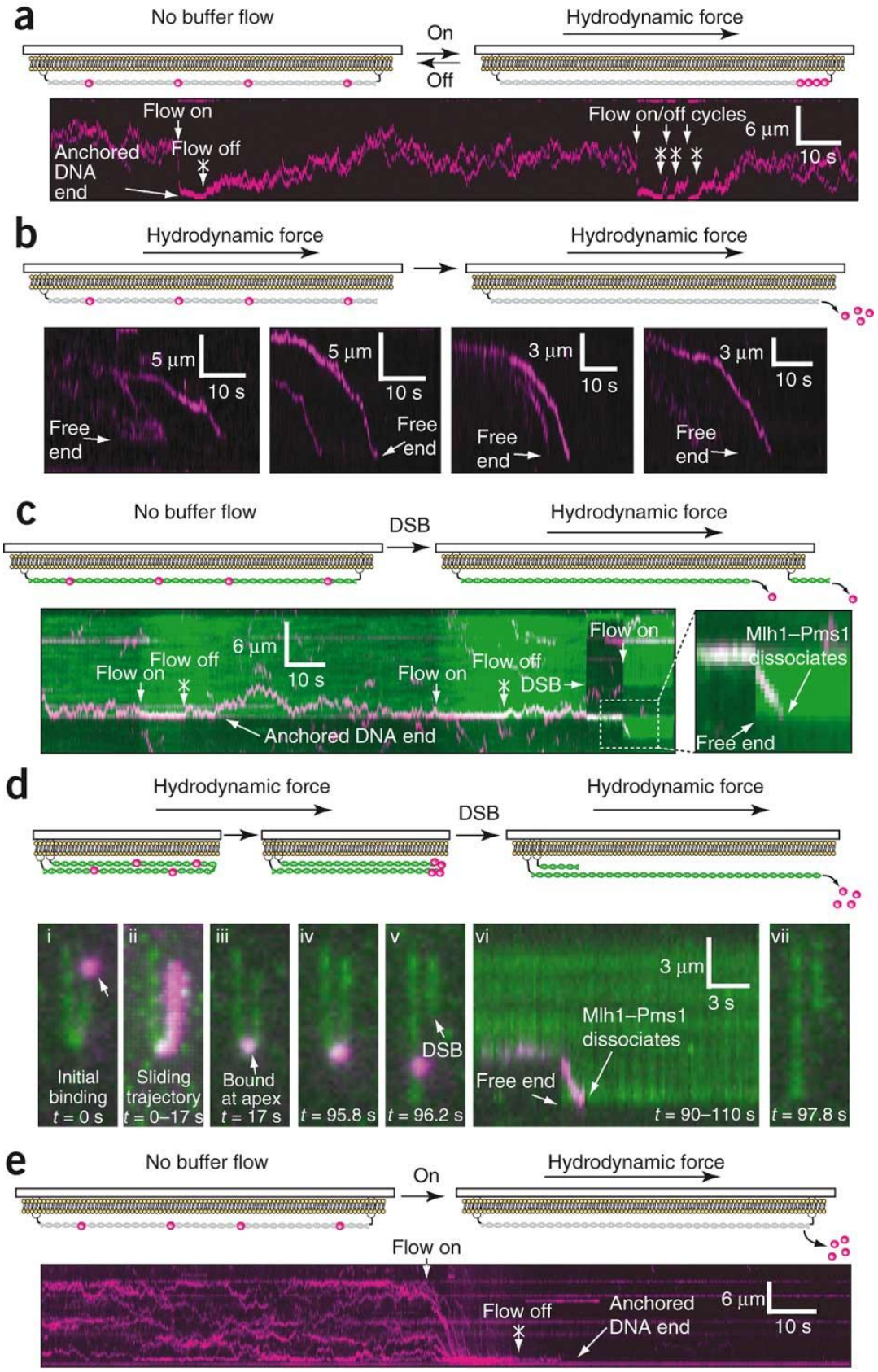
Figure 2.3. Quantitative analysis of Mlh1–Pms1 diffusion.
(A) Examples of MSD plots generated from tracking the motion of Mlh1–Pms1. **(B)** Diffusion coefficients derived from the MSD plots and categorized according into the different nucleotide conditions used for the measurements. Diamonds (♦), the mean values; the mean ± s.d. are color coded (graph and inset; $N \geq 25$ for all reported diffusion coefficients). The cumulative diffusion coefficients for Mlh1–Pms1 and the previously measured values for Msh2–Msh6 are shown together for comparison (Gorman *et al.* 2007), and the upper boundaries for the theoretical diffusion coefficients based on models either with or without a rotational component are also shown (Gorman *et al.* 2007; Bagchi *et al.* 2008). **(C)** Diffusion coefficients for Mlh1–Pms1 determined at different concentrations of NaCl.

mechanisms for diffusive motion along DNA include hopping, jumping, sliding or intersegmental transfer. The structure of Mlh1–Pms1 (Figure 2.1A) also suggested a possible 'stepping' mechanism, which is virtually identical to hopping, with the N- and/or C-terminal domains (NTD and CTD, respectively) acting as DNA-binding domains that independently hop while connected by flexible linkers (Guarne *et al.* 2001; Guarne *et al.* 2004; Kosinski *et al.* 2005; Sacho *et al.* 2008). Jumping would yield punctate kymograms as a consequence of repeated dissociation and rebinding events and cannot account for the continuous motion that predominated the diffusion trajectories; the stretched DNA configuration makes intersegmental transfer involving DNA-looping unlikely; and the 38-fold increase (Student's *t*-test, $P < 0.0001$) in the diffusion coefficient measured over a range of salt concentrations argues against sliding ($D_{1D} = 0.026 \pm 0.017$ μm^2/sec versus 0.99 ± 0.411 μm^2/sec, at 25 and 200 mM NaCl, respectively; Figure 2.3C and Appendix, Supplementary Figure 2.8) but is consistent with a hopping and/or stepping mechanism (see Appendix, Supplementary Discussion) (Blainey *et al.* 2006; Komazin-Meredith *et al.* 2008). In contrast to Mlh1–Pms1, we have previously shown that the diffusion coefficient of Msh2–Msh6 does not vary over the same range of NaCl concentrations (see Supplementary Figure 3 from (Gorman *et al.* 2007)), which is most consistent with a sliding mechanism (Gorman *et al.* 2007). We conclude that, although Msh2–Msh6 and Mlh1–Pms1 both travel along DNA via 1D diffusion, they do so using different mechanisms: Msh2–Msh6 slides while in continuous contact with the phosphate backbone, whereas Mlh1–Pms1 hops or steps as it moves back and forth along DNA.

Mlh1–Pms1 has properties consistent with ring-like architecture. The Mlh1–Pms1

heterodimer is maintained through protein-protein interactions between the CTDs, and the NTDs and CTDs are separated by very long linker arms (Hall *et al.* 2003; Sacho *et al.* 2008) (Figure 2.1A). It has been previously hypothesized that this structural organization allows Mlh1–Pms1 and related proteins to adopt a ring-like architecture, which would enable them to encircle DNA (Guarne *et al.* 2004) (Figure 2.1A). This type of topological binding mechanism leads to several specific experimentally testable predictions: (i) dissociation from DNA should occur preferentially from the free ends of linear DNA molecules; (ii) protein dissociation should be prevented if the DNA ends are sterically occluded; (iii) dissociation should be less prevalent from internal positions; and (iv) an intact heterodimer would be necessary for stable DNA-binding activity and end-dependent dissociation. We first asked whether Mlh1–Pms1 preferentially dissociated from DNA ends (Figure 2.4A-D). When hydrodynamic force (~100 fN) was used to push Mlh1–Pms1, most complexes (>95%) did not dissociate upon encountering anchored (that is, sterically blocked) DNA ends (Figure 2.4A; N_{dis}/N_{tot} = 1/23 (dissociated / total pushed to DNA ends)), nor did Mlh1–Pms1 dissociate from the apex of looped DNA (Figure 2.4D N_{dis}/N_{tot} = 0/4). In contrast, Mlh1–Pms1 immediately dissociated from free ends of 'single-tethered' DNA (Figure 2.4B; N_{dis}/N_{tot} = 880/1,000) and from free ends of photochemically induced double-stranded breaks (DSBs; Figure 2.4C,D; N_{dis}/N_{tot}= 14/14). Mlh1 alone can exist as monomers or dimers (K_d = 3.14 ± 0.19 μM), but the Mlh1-NTDs do not self associate (Hall *et al.* 2003). Mlh1 alone could bind DNA (Figure 2.4E), with a diffusion coefficient 6.9-fold greater ($P < 0.0001$) than that of Mlh1–Pms1 under the same conditions (D_{1D} = 0.137 ± 0.127 μm^2/sec, N = 25, versus 0.020 ± 0.023 μm^2/sec, N = 25, respectively, at 50 mM NaCl; we did not detect Mlh1 binding at higher ionic strengths), indicating that Pms1 was not essential for binding or diffusion.

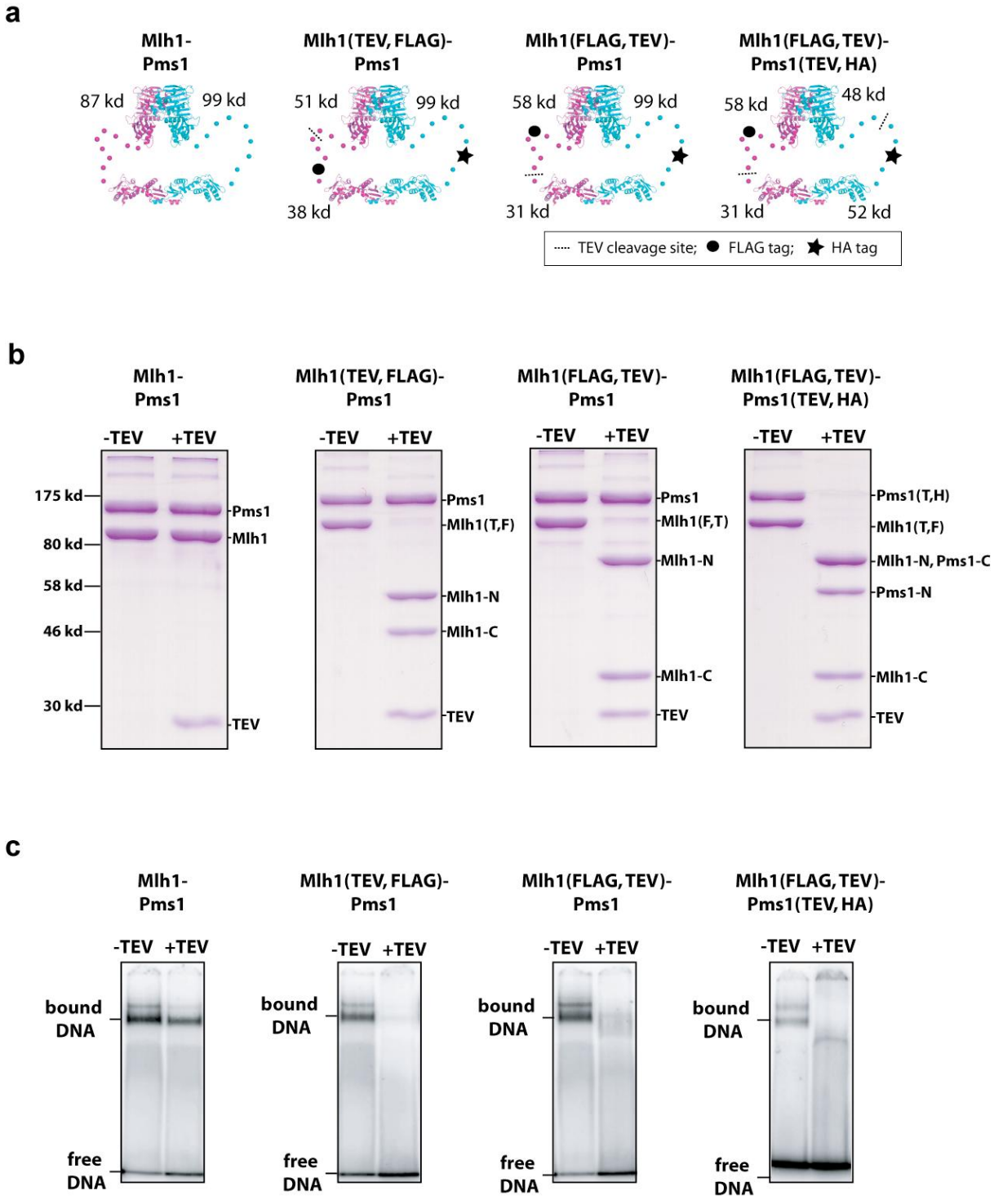
Figure 2.4. End-dependent dissociation of Mlh1–Pms1 from DNA.
(A) Kymogram of Mlh1–Pms1 (magenta) diffusing on a DNA molecule (unlabeled) anchored by both ends to the flow-cell surface. Flow is cycled on and off as indicated. When flow is on, Mlh1–Pms1 is pushed to the downstream anchored end of the DNA but does not dissociate. **(B)** Kymograms of Mlh1–Pms1 dissociating from the free blunt ends (generated by SfoI digest) of single-tethered DNA molecules. **(C)** Mlh1–Pms1 (magenta) was bound to DNA (green) stained with YOYO1 and was pushed repeatedly to the end of the molecule (as indicated) to verify that it did not dissociate. A DSB was introduced by laser illumination in the absence of flow. Upon breaking, the DNA retracts from the surface, as indicated by the sudden disappearance of the green signal. Flow was resumed to extend the broken DNA and push Mlh1–Pms1 toward the free end of the molecule. Mlh1–Pms1 immediately dissociates upon encountering the free DNA end (see inset). **(D)** Mlh1–Pms1 was bound to a looped DNA molecule (i) and slid along the arc formed by the DNA until stopping at the loop apex (ii). The protein remained at the DNA apex (iii and iv) but continued sliding down the DNA upon induction of a DSB (v), and the protein immediately dissociated upon reaching the newly generated free end (vi), leaving behind the naked DNA (vii). **(E)** Kymogram of Mlh1 bound to a double-tethered DNA molecule. In the absence of flow the proteins diffuse rapidly along the DNA, but when flow is applied, they are pushed to the end of the DNA and rapidly dissociate. Experiments in **(A**–**E)** were collected from isolated DNA molecules, as described (Gorman *et al.* 2007). The DNA in **(A** and **B)** was located with YOYO1, but the dye was removed before data acquisition to avoid unintentional photocleavage, and the experiments were conducted at 150 mM NaCl. The experiments in **(C**, **D** and **E)** were conducted at 50 mM NaCl. In all cases, identical results were obtained ±1 mM ATP.


a
No buffer flow
On
Off
Hydrodynamic force
Flow on
Flow off
Anchored DNA end
Flow on/off cycles
6 μm
10 s
b
Hydrodynamic force
Hydrodynamic force
5 μm
10 s
Free end
5 μm
10 s
Free end
3 μm
10 s
Free end
3 μm
10 s
Free end
c
No buffer flow
DSB
Hydrodynamic force
6 μm
10 s
Flow on
Flow off
Flow on
Flow off
Flow on
DSB
Anchored DNA end
Mlh1–Pms1 dissociates
Free end
d
Hydrodynamic force
DSB
Hydrodynamic force
i
Initial binding t = 0 s
ii
Sliding trajectory t = 0–17 s
iii
Bound at apex t = 17 s
iv
t = 95.8 s
v
DSB
t = 96.2 s
vi
3 μm
3 s
Mlh1–Pms1 dissociates
Free end
t = 90–110 s
vii
t = 97.8 s
e
No buffer flow
On
Hydrodynamic force
Flow on
Flow off
Anchored DNA end
6 μm
10 s

However, when flow was applied, Mlh1 moved rapidly down the DNA, and >80% dissociated from the anchored DNA ends (Figure 2.4E N_{dis}/N_{tot} = 285/350). This finding was strikingly different from results with the intact heterodimer, indicating that the presence of Pms1 was necessary to observe end-dependent dissociation. Finally, we engineered TEV cleavage sites into the linker of Mlh1 and Pms1, and proteolytic cleavage of one or both linker arms abolished detectable DNA-binding activity (Figure 2.5), highlighting the importance of the linker arms for DNA binding. We conclude that the formation of an intact Mlh1–Pms1 heterodimer stabilizes the DNA-bound complex and that the heterodimer preferentially dissociates from DNA ends. These experimental findings are all consistent with predictions for the previously proposed mechanism whereby Mlh1–Pms1 can adopt a ring-like architecture that wraps around DNA, although we are careful to note that we do not yet know the structural details of the wrapped complex.

Mlh1–Pms1 complexes can bypass one another while traveling along DNA. We have previously shown that Msh2–Msh6 complexes traveling on the same molecule cannot pass one another, arguing that the proteins maintain continuous close contact with the DNA, which is consistent with a sliding mechanism (Gorman *et al.* 2007). In contrast, two-color labeling experiments revealed that Mlh1–Pms1 complexes could bypass one another as they traveled along the same DNA molecule (Appendix, Supplementary Figure 2.9), which is only consistent with a hopping/stepping mechanism wherein the individual hops or steps span distances comparable to or greater than the dimensions of the QD-tagged proteins. Closed ring-like architecture is difficult to reconcile with the observed protein bypass and would require two Mlh1–Pms1 complexes to thread through one another as they moved along the

Figure 2.5. TEV cleavage of Mlh1-Pms1 linker arms disrupts DNA-binding.
(A) Schematic overview of different TEV-containing Mlh1-Pms1 constructs. **(B)** Coomassie-stained SDS-PAGE showing specificity of TEV cleavage for each of the different constructs. **(C)** Gel shift assays using a ^{32}P-labeled oligonucleotide substrate ±TEV cleavage. All of the proteins bind DNA before TEV cleavage, but treatment with TEV protease reduces or eliminates DNA binding activity in the bulk assay. Similarly, all of these protein constructs bound and diffused on DNA in the TIRFM assays, but we could detect no DNA binding activity in the TIRFM assays after TEV cleavage of the linker arms (data not shown).


a
Mlh1-Pms1
87 kd
99 kd
Mlh1(TEV, FLAG)-Pms1
51 kd
99 kd
38 kd
Mlh1(FLAG, TEV)-Pms1
58 kd
99 kd
31 kd
Mlh1(FLAG, TEV)-Pms1(TEV, HA)
58 kd
48 kd
31 kd
52 kd
TEV cleavage site; FLAG tag; HA tag
b
Mlh1-Pms1
-TEV +TEV
175 kd
80 kd
58 kd
46 kd
30 kd
Pms1
Mlh1
TEV
Mlh1(TEV, FLAG)-Pms1
Pms1
Mlh1(T,F)
Mlh1-N
Mlh1-C
TEV
Mlh1(FLAG, TEV)-Pms1
Pms1
Mlh1(F,T)
Mlh1-N
Mlh1-C
TEV
Mlh1(FLAG, TEV)-Pms1(TEV, HA)
Pms1(T,H)
Mlh1(T,F)
Mlh1-N, Pms1-C
Pms1-N
Mlh1-C
TEV
c
Mlh1-Pms1
-TEV +TEV
bound DNA
free DNA
Mlh1(TEV, FLAG)-Pms1
Mlh1(FLAG, TEV)-Pms1
Mlh1(FLAG, TEV)-Pms1(TEV, HA)

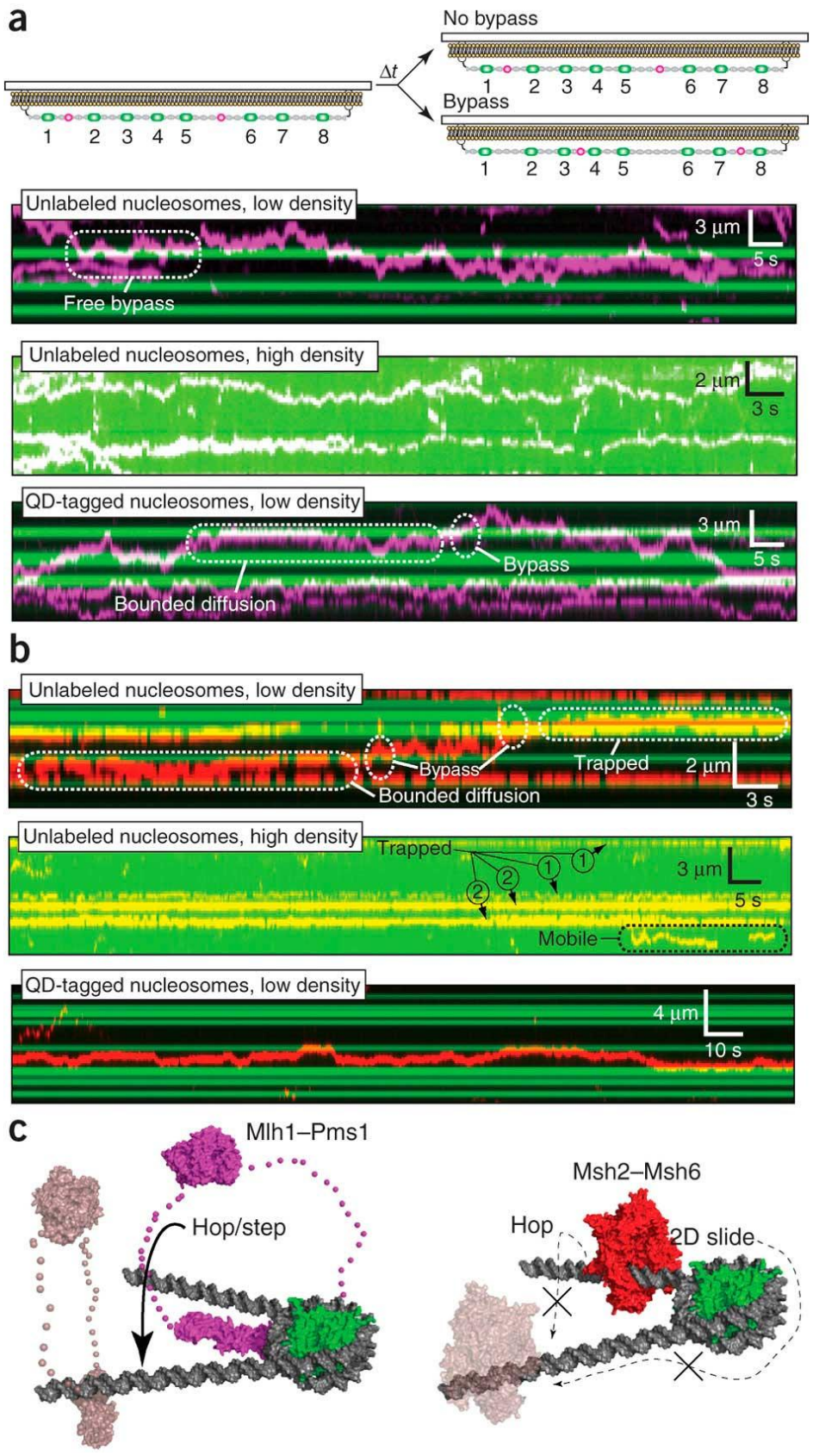
DNA. A threading mechanism specifically predicts that Mlh1–Pms1 would be unable to bypass obstacles larger than the internal diameter of the large central pore formed by the protein complex. Alternatively, bypass could also be accomplished through transient ring opening, whereupon the proteins could simply step past one another in an open configuration. This type of open stepping mechanism predicts that Mlh1–Pms1 would be capable of bypassing obstacles larger than the internal diameter of the protein ring. Given the combined length of the Mlh1–Pms1 linker arms (51.7 ± 14.6 nm) (Sacho *et al.* 2008), the corresponding maximal diameter of the central pore would be 16.5 ± 4.6 nm in diameter, which is too small to accommodate the passage of a QD (~20 nm in diameter), ruling out a threading mechanism for obstacle bypass. We conclude that Mlh1–Pms1 most likely bypasses obstacles by stepping over them in an open ring configuration, implying that the protein is capable of transitioning back and forth between an open and closed conformation.

Mlh1–Pms1 can traverse nucleosomes during 1D diffusion. The finding that Mlh1–Pms1 complexes could bypass one another suggested that these proteins might be able to undergo 1D diffusion on crowded DNA substrates, similar to what would be found in an *in vivo* environment. Therefore, we next asked whether MMR proteins could traverse nucleosomes, which are anticipated to be the most abundant obstacles encountered in eukaryotes. For these experiments, we deposited unlabeled, recombinant nucleosomes onto the DNA substrates by salt dialysis at a ratio of either ~5–10 or ~80–100 nucleosomes per DNA molecule (Visnapuu and Greene 2009). We performed the Mlh1–Pms1 diffusion measurements as described above, and we then located the nucleosomes by labeling them with QDs after the diffusion measurements were completed. Mlh1–Pms1 still diffused on nucleosome-bound DNA (D_{1D} =

0.027 ± 0.021 μm^2/sec, $N = 26$) and repeatedly bypassed unlabeled nucleosomes (~10 nm in diameter), showing no evident boundary effects upon colliding with single nucleosomes (Figure 2.6A, upper panel; $N > 1{,}000$ Mlh1–Pms1 complexes, each giving rise to multiple bypass events). Mlh1–Pms1 also moved freely along DNA bound by up to ~80–100 unlabeled nucleosomes (Figure 2.6A, middle panel; $D_{1D} = 0.034 \pm 0.018$ μm^2/sec, $N = 25$), providing an unequivocal demonstration that nucleosomes do not prevent 1D diffusion of Mlh1–Pms1. We conclude Mlh1–Pms1 can travel along a simple chromatin lattice by 1D diffusion while bypassing protein obstacles at it travels along the DNA. As indicated above, for all of these experiments, we labeled the nucleosomes only after making the diffusion measurements to ensure that the large QDs would not interfere with Mlh1–Pms1 movement. However, Mlh1–Pms1 could also bypass QD-labeled nucleosomes ($N = 63$ Mlh1–Pms1 complexes, each yielding multiple bypass events), although in this case, Mlh1–Pms1 showed characteristics of bounded diffusion upon colliding with the QD-labeled nucleosomes, with the large QD-labeled nucleosomes acting as semi-penetrable barriers (Figure 2.6A, lower panel). Given the large diameter of the QD-labeled nucleosome (≥30 nm) compared to the size of Mlh1–Pms1, we conclude that nucleosome bypass must occur via a stepping mechanism whereby the protein transiently adopts an open ring configuration. These results provide the first experimental demonstration that a protein undergoing 1D diffusion can circumnavigate protein obstacles that lie in its path.

Diffusion of Msh2–Msh6 is restricted by nucleosomes. In striking contrast to Mlh1–Pms1, the movement of Msh2–Msh6 past unlabeled nucleosomes was highly restricted, showing characteristics of bounded diffusion with nucleosomes acting as semi-penetrable barriers, and

Figure 2.6. Diffusion of Mlh1–Pms1 and Msh2–Msh6 along nucleosome-bound DNA. **(A)** Mlh1–Pms1 (magenta) is shown diffusing along a DNA bound by recombinant nucleosomes (green); white, regions of overlapping signal. The nucleosomes were labeled with QDs either after conducting the Mlh1–Pms1 diffusion experiment (upper and middle panels) or before the addition of Mlh1–Pms1 (lower panel). Examples of bypass and bounded diffusion are highlighted; identical results were obtained ± ATP. **(B)** Msh2–Msh6 (red) is shown diffusing on a DNA molecule with unlabeled nucleosomes (green); yellow, regions of overlap. Upper and middle panels, results with unlabeled nucleosomes. The number designations (either 1 or 2) in the middle panel indicate number of QD-tagged Msh2–Msh6 molecules trapped in each region of the kymogram (6 total). Lower panel, Msh2–Msh6 colliding with QD-labeled nucleosomes. **(C)** Illustrates how the structures of Mlh1–Pms1 (left) and Msh2–Msh6 (right) may influence nucleosomal encounters (Luger *et al.* 1997; Guarne *et al.* 2001; Guarne *et al.* 2004; Kosinski *et al.* 2005; Warren *et al.* 2007). The molecules are drawn to scale. The trajectory of the DNA leaving the nucleosome surface has been modified for illustrative purposes. Mlh1–Pms1 steps over the nucleosome (solid arrow). Nucleosome bypass by Msh2–Msh6 might occur by occasional hopping or two-dimensional (2D) sliding (dashed arrows) (Kampmann 2005), but neither mechanism allows free mobility on a nucleosomal array.


a
No bypass
Δt
1 2 3 4 5 6 7 8
Bypass
1 2 3 4 5 6 7 8
1 2 3 4 5 6 7 8
Unlabeled nucleosomes, low density
3 μm
5 s
Free bypass
Unlabeled nucleosomes, high density
2 μm
3 s
QD-tagged nucleosomes, low density
3 μm
5 s
Bypass
Bounded diffusion
b
Unlabeled nucleosomes, low density
Bypass
Trapped
2 μm
3 s
Bounded diffusion
Unlabeled nucleosomes, high density
Trapped
3 μm
5 s
Mobile
QD-tagged nucleosomes, low density
4 μm
10 s
c
Mlh1–Pms1
Hop/step
Msh2–Msh6
Hop
2D slide

Msh2–Msh6 typically became trapped between nucleosomes (Figure 2.6B, upper and middle panels). On higher-density nucleosome arrays (~80–100 nucleosomes per 48.5 kb DNA substrate) most molecules of Msh2–Msh6 were immobile or oscillated within tightly confined regions (N = 964 / 1000) and showed little evidence of free 1D diffusion within our detection limits ($D_{1D} \leq 1 \times 10^{-4}$ μm^2/sec). A small subpopulation of Msh2–Msh6 remained mobile on the high-density arrays (N = 36 / 1000; 3.6%), suggesting that they were bound in an alternate conformation. In further contrast to Mlh1–Pms1, Msh2–Msh6 never bypassed QD-tagged nucleosomes (Figure 2.6B, lower panel), indicating that large obstacles (≥30 nm in diameter) present insurmountable barriers, which is fully consistent with expectations based on the structure of Msh2–Msh6, which wraps around DNA, making intimate contacts with the phosphate backbone over nearly 1.5 turns of helix (Warren *et al.* 2007), and is also consistent with a continuous sliding mechanism that does not involve extensive hopping (Gorman *et al.* 2007). Rare nucleosome bypass by Msh2–Msh6 might occur through occasional hopping events or through limited excursions into two-dimensional sliding, where Msh2–Msh6 maintains contact with the DNA without tracking the helical pitch of the phosphate backbone (Kampmann 2005; Gorman and Greene 2008) (Figure 2.6C). In either case, the mechanism does not permit efficient mobility of Msh2–Msh6 along the higher-density nucleosome arrays. This conclusion agrees with bulk biochemical studies showing that nucleosomes or other stationary obstacles can trap Msh2–Msh6 (or its homologs) on DNA (Gradia *et al.* 1999; Mendillo *et al.* 2005; Pluciennik and Modrich 2007; Li *et al.* 2009).

Discussion

Intranuclear trafficking of all DNA-binding proteins is governed by facilitated diffusion. Theoretical descriptions and bulk measurements of facilitated diffusion have long been reported in the literature, beginning with the classical studies of l*ac* repressor (Riggs *et al.* 1970; Berg *et al.* 1981; Winter *et al.* 1981; Winter and von Hippel 1981) and more recently with NMR experiments of transcription factors (Iwahara and Clore 2006; Iwahara *et al.* 2006; Doucleff and Clore 2008), but direct measurements of diffusion have only recently become possible through the development of new single-molecule techniques (Blainey *et al.* 2006; Gorman *et al.* 2007; Gorman and Greene 2008; Liu *et al.* 2008; Tafvizi *et al.* 2008; Roy *et al.* 2009). Together, these studies support an emerging consensus that many DNA-binding proteins can travel long distances along DNA by 1D diffusion *in vitro*. However, the validity of this conclusion with respect to physiological settings remains unclear despite years of experimental and theoretical efforts, specifically because it remains unknown whether or how 1D diffusion can occur in the presence of nucleosomes and other nucleoprotein structures (von Hippel and Berg 1989; Halford and Szczelkun 2002; Kampmann 2005; Gorski *et al.* 2006; Hager *et al.* 2009; Li *et al.* 2009; Mirny *et al.* 2009). Here we sought to resolve this issue by using single-molecule imaging, nanofabricated curtains of double-tethered DNA molecules, and MMR proteins as model systems for facilitated diffusion.

We have shown that both Msh2–Msh6 and Mlh1–Pms1 can diffuse in 1D along DNA but do so using very different mechanisms. Mlh1–Pms1 hops or steps along the DNA, but Msh2–Msh6 moves predominantly by sliding along the DNA while remaining in continuous close contact with the phosphate backbone. The functional consequences of these mechanistic

differences are that Mlh1–Pms1 can readily traverse nucleosomes and travel along chromatin, whereas Msh2–Msh6 cannot. These results provide an unambiguous demonstration that 1D diffusion can occur on crowded DNA substrates in the presence of protein obstacles and that the ability to bypass obstacles is dependent on the how the protein in question diffuses along DNA. We anticipate that these behaviors shown by Mlh1–Pms1 and Msh2–Msh6 in response to collisions with nucleosomes will reflect general mechanistic attributes of their respective modes of 1D diffusion, which will apply in principle to any proteins that diffuse on DNA (for example, DNA repair proteins, transcription factors, etc.). Proteins that track the phosphate backbone while sliding along DNA will experience a barrier upon encountering obstacles. Therefore either the protein must disengage the DNA and enter a two- or three-dimensional mode of diffusion to continue searching for targets or the DNA must be cleared of obstacles beforehand to allow unhindered access to the DNA (see below). In contrast, proteins that do not track the backbone can traverse obstacles without experiencing boundary effects.

The different modes of diffusion found for Msh2–Msh6 and Mlh1–Pms1 also impose specific constraints on the mechanisms of MMR. Msh2–Msh6 is the first to arrive at lesions, and helps identify nearby signals differentiating the parental and nascent DNA strands. Many models for strand discrimination invoke 1D movement of Msh2–Msh6 along DNA, and even transient loss of contact with the DNA during this second phase of the reaction could compromise repair (Kunkel and Erie 2005; Jiricny 2006; Modrich 2006). Nucleosomes, or other DNA-binding proteins, have the potential to thwart Msh2–Msh6, and a single nucleosome deposited near a lesion could render it irreparable, suggesting that regions in need of repair must be kept free from obstacles. Replication forks disrupt nucleosomes, leaving stretches of naked DNA in their wake (Groth *et al.* 2007). Although it is speculative, if Msh2–

Msh6 were restricted to the region behind the fork, possibly through direct association with PCNA, then it would be free to scan newly replicated, naked DNA (Umar *et al.* 1996; Gorman *et al.* 2007; Kolodner *et al.* 2007). Mlh1–Pms1 is thought to arrive later than Msh2–Msh6 (Kunkel and Erie 2005; Jiricny 2006), implying that it must survey the entire genome for lesion-bound Msh2–Msh6 without the benefit of confined searches in regions already cleared by a replication fork. The ability of Mlh1–Pms1 to hop or step along DNA and to freely traverse nucleosomes ensures that it could efficiently bypass stationary obstacles while searching the genome for its binding targets.

Materials and Methods

Mlh1–Pms1 expression, purification and characterization. Mlh1–Pms1 was expressed and purified using pMH1 (*GAL1-MLH1-VMA-CBD*, 2μ, *TRP1*) and pMH8 (*GAL10-PMS1*, 2μ, *LEU2*) vectors transformed into the *S. cerevisiae* strain BJ2168 (Hall *et al.* 2001). Mlh1–Pms1 complexes containing Flag, hemagglutinin and/or TEV tags were purified from BJ2168 containing the relevant pMH1 and pMH8 derivatives described below. Mlh1–Pms1 constructs were functional for mismatch repair *in vivo* (see below).

Mlh1-Pms1 cloning and characterization. Plasmids containing tagged *MLH1* under the native *MLH1* promoter were created by overlap-extension PCR as derivatives of pEAA213 (Heck *et al.* 2006): pEAA373 contains a Flag-tag (underlined) flanked on either side by three alanines (AAADYKDDDKAAA) and inserted immediately after amino acid 448T of *MLH1*; pEAA516 contains a TEV site (underlined) flanked on either side by three alanines

(AAAENLYFQSAAA) inserted immediately after amino acid 448T and the Flag-tag inserted immediately after amino acid 499Y of *MLH1*. For expression, the tagged *MLH1* constructs were sub-cloned into pMH1 (a generous gift from Dr. Tom Kunkel (NIEHS), and purified as described (Hall and Kunkel 2001). Sub-cloning from pEAA373 and pEAA516 into pMH1 created pEAE267 and pEAE295, respectively. For tagged *PMS1*, pEAE296 was created as a derivative of pMH8 (gift from Dr. Tom Kunkel, NIEHS) by overlap-extension PCR and included a HA-tag (underlined) flanked by three alanine residues (AAAYPYDVPDYAAAA) inserted after amino acid D565 of *PMS1*. For complementation, the epitope-tag from pEAE296 was sub-cloned into pEAA238 (Heck *et al.* 2006), which has *PMS1* under its native promoter, and then cloned into pEAA248, which is the same as pEAA238 but contains the *URA3* selectable marker, to create pEAA517. All clones were sequenced (Cornell BioResource Center), and additional details on vector construction will be provided upon request.

MLH1 constructs were tested *in vivo* for the *mlh1Δ* mutator phenotype in *ARS-CEN LEU2* vectors containing MLH1 expressed from its native promoter. The HA-tagged PMS1 was tested *in vivo* for the *pms1*Δ mutator phenotype (Heck *et al.* 2006). To test complementation, the semiquantitative canavanine resistance assay was used to measure mutation rates in the S288c strains EAY874 (*MATα, leu2-3,112, trp1-289, ura3::argD, cyhS, mlh1*Δ*::KanMX4*) and EAY1087 (*MATa, ura3-52, leu2*Δ*1, trp1*Δ*63, his3*Δ*200, lys2*Δ*202, pms1*Δ*::KanMX4*). EAY874 strains containing pRS415 (mutant control), pEAA213 (wild-type control) or pEAA373, pEAA515 and pEAA516 (epitope-tagged *MLH1*) were streaked on leucine dropout plates to obtain single colonies. EAY310 strains containing pRS416 (mutant control), pEAA248 (wildtype control) and pEAA517 (*HA-PMS1*) were streaked onto

uracil dropout plates to obtain single colonies. 35 independent colonies from each strain were patched onto appropriate dropout plates containing L-canavanine (60 mg/L) and incubated for 3 days at 30°C. The number of canavanine-resistant papillations in each patch was counted and the median number for each strain was used for comparison. All tested derivatives of MLH1 and PMS1 conferred mutation frequencies indistinguishable from wild-type in contrast to the 10-fold higher frequency found in the corresponding null strains carrying an empty vector.

Ensemble characterization of Mlh-Pms1 DNA-binding activity. Protein concentrations were determined by Bradford assay using BSA as a standard. Gel mobility shift assays with oligonucleotide substrates were performed as described (Kijas *et al.* 2003). Binding reactions were performed at room temperature (RT) for 5 minutes in 15 µl reactions containing 60 nM (5'-^{32}P)-end labeled 40-bp substrate, 120 nM Mlh1-Pms1, 25 mM Hepes [pH, 7.6], 40 µg/ml BSA, 1mM DTT, 50 mM NaCl, and 8% (w/v) sucrose. In super-shift experiments, either 0.65 µg of Flag antibody (Sigma Cat. No. F3165) or 3.25 µg of HA antibody (Sigma Cat. No. H3663) were preincubated with Mlh1-Pms1 on ice for 30 min prior to the addition of DNA substrates. Samples were loaded on 4% (w/v) non-denaturing polyacrylamide gels containing 0.5X TBE and electrophoresed at 130 V for 1 hour at RT. Gels were dried on 3MM Whatman paper and visualized by PhosphorImaging. Analysis was done using ImageJ. The 40-bp substrate was created by annealing S1 (5'dACCGAATTCTGACTTGCTAGGACATCTTTGCCCACGTTGA) and S2 (5'dTCAACGTGGGCAAAGATGTCCTAGCAAGTCAGAATTCGGT) (Integrated DNA Technologies; (Surtees and Alani 2006).

Nitrocellulose filter binding assays were performed as described (Chi and Kolodner 1994). Briefly, nitrocellulose filters (BA85, 0.45 μm, 25 mm, Whatman Schleicher & Schuell) were presoaked in 0.5 M KOH for 20 minutes, rinsed thoroughly with sterile deionized distilled water, washed once with reaction buffer (25 mM Hepes [pH 7.6], 0.01 mM EDTA) and stored in same buffer prior to use. The 3 kb plasmid (pEAO242) was linearized with NcoI followed by treatment with calf intestinal phosphatase (CIP; New England Biolabs). CIP-treated DNA was purified (QIAquick; Qiagen) and end-labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP, followed by heat inactivation and removal of unincorporated nucleotides, as described by the manufacturer (New England Biolabs). Binding reactions (30 μl) were performed in buffer supplemented with 1 mM DTT and 40 μg/ml BSA without EDTA, and included 100 pM of 5'^{32}P-labeled 3 kb linear plasmid. Mlh1-Pms1 was incubated with 40 nM IgG or IgG coupled QDs (as indicated) on ice for 30 minutes. Binding reactions were incubated at RT for 10 minutes, and filtration was performed on a Hoefer filter manifold (FH225V, Hoefer Scientific). Binding reactions were added to 2.5 ml of ice-cold reaction buffer overlaying nitrocellulose filters and passed through with a flow rate of ~2.5 ml/min. Dried filters were placed in 5 ml of Ecoscint scintillant (National Diagnostics) and the bound radioactivity measured in a Beckman LS 5000 scintillation counter. Background was determined in reactions without protein and was typically around 1-5% of total radioactivity. Percent DNA binding was determined by dividing the background subtracted count for each filter by the total radioactivity per reaction.

TEV cleavage assays. 3 μg of Mlh1-Pms1 was incubated with 0.03 μg of TEV protease in 15 μl reactions containing 25 mM Hepes [pH 7.6], 1 mM DTT, and 40 μg/ml BSA. TEV

protease was a generous gift from Dr. Ailong Ke (Cornell). TEV cleaved proteins were assayed for DNA binding activity using gel shift assays, as described above (Figure 2.5). To confirm TEV cleavage, samples were incubated at 30°C for 30 minutes, after which 7.5 μl of 3X SDS-loading buffer (0.195 M Tris [pH 6.8], 30% glycerol, 3% β-mercaptoethanol, 6% SDS) was added to each and samples were boiled for 3 minutes. Samples were analyzed by 10% SDS-PAGE and stained with Coomassie blue (Figure 2.5).

Single-molecule imaging. The TIRFM system, particle tracking and data analysis have been described previously (Gorman *et al.* 2007). Unless otherwise indicated, data were collected using DNA molecules anchored with nanofabricated patterns made by electron-beam lithography (Gorman *et al.* 2010). QDs were prepared using a protocol that yields 0.076 ± 0.014 epitope binding sites per QD (that is, ~1/13 QDs has half of a functional IgG) (Pathak *et al.* 2007). Assuming the conjugation reaction is a Poisson process, the probability of a QD having two or three epitope binding sites is $P = 0.0027$ and $P = 7.02 \times 10^{-5}$, respectively. Flag-tagged Mlh1–Pms1 (30–250 nM) was mixed with two-fold molar excess anti-Flag QDs in buffer containing 40 mM Tris-HCl (pH 7.7), 150 mM NaCl, ±1 mM $MgCl_2$, 1 mM DTT, and 0.4 mg/ml BSA and was incubated for 15–20 min on ice. Reactions were diluted ten-fold (3–25 nM Mlh1–Pms1) before injection. Cross-linking assays verified that Mlh1–Pms1 did not dissociate to monomers under dilute conditions (data not shown). TIRFM experiments were done using 40 mM Tris-HCl (pH 7.7), 150 mM NaCl (unless otherwise indicated), ±1 mM $MgCl_2$, 1 mM DTT, 0.4 mg/ml BSA, 140 mM β-mercaptoethanol, ±1 mM nucleotide (ADP, ATP or ATPγS, as indicated). Labeled proteins were injected into the sample chamber, unbound proteins were quickly flushed away and flow was terminated before data acquisition.

YOYO1 (0.5 nM; Invitrogen) and an O_2 scavenging system (glucose oxidase (34 units/ml), catalase (520 units/ml), and 1% (w/v) glucose) was included in reactions requiring fluorescent DNA. Bulk experiments verified that YOYO1 and the O_2 scavenging system did not affect Mlh1–Pms1 DNA-binding activity (not shown). Recombinant histones were purified from *E. coli*, reconstituted into octamers and deposited on the DNA as described (Visnapuu and Greene 2009).

Construction of DNA substrates for TIRFM. DNA substrates were made by ligating oligonucleotides to the 12-nucleotide overhangs at the end of the λ-DNA. Oligonucleotides were purchased from Operon Technologies and gel purified prior to use. Ligation mixes (1 ml total volume) contained 4 nM λ-DNA (Invitrogen), 1 μM biotinylated oligonucleotide (5' pAGGTCGCCGCCC[BioTEG]-3'), 1 μM DIG (digoxigenin) or FITC (fluorescein isothiocyanate) labeled oligonucleotide (5'-pGGG CGG CGA CCT[DIG]-3' or 5'-pGGGCGGCGACCT[FITC]-3'), and 1X ligase buffer (New England Biolabs). The reaction mix was warmed to 65°C for 10 minutes and then cooled slowly to RT. After cooling, ligase was added (T4 DNA ligase (400 units/μl) or Taq ligase (40 units/μl; New England Biolabs) and the mixture was incubated overnight at 42°C. Reactions performed with T4 ligase were heat inactivated at 65°C for 10 minutes, and ligated DNA products were purified over a Sephacryl S200HR column (GE Healthcare) run in 10 mM Tris-HCl [pH 7.8], 1 mM EDTA, plus 150 mM NaCl. Purified DNA was stored at –20°C.

Quantum Dots. QDs were prepared by following the manufacturer's recommended protocol, as described in (Gorman *et al.* 2007) and (Pathak *et al.* 2007) (Qdot® 585 Antibody

Conjugation Kit, Cat. No. Q22011MP, and Qdot® 705 Antibody Conjugation Kit, Cat. No. Q22061MP; Invitrogen). In brief, QDs (4 μM) were activated with 1 mM SMCC (succinimidyl 4-[Nmaleimidomethyl] cyclohexane-1-carboxylate) for 1 hour at RT, yielding maleimide functional groups on the QD surface. Antibodies (1 mg/ml in PBS; Sigma Cat. No. H3663 and F3165, for anti-HA and anti-Flag, respectively) were reduced with 20 mM DTT for 30 minutes at RT, which cleaves disulfide bonds between heavy chains generating free thiols that can be coupled the maleimide-QDs. Reduced IgG was purified on a NAP-5 column and then mixed with the maleimide activated QDs for 1 hour at RT. Reactions were quenched with 100 μM β-mercaptoethanol for 30 minutes at RT, and the resulting antibody-QD conjugates were purified over a Superdex 200 10/300 GL gel filtration column (GE Healthcare) run in PBS to remove unreacted IgG. Purified conjugates were stored in PBS [pH 7.4] plus 0.1 mg/ml acetylated BSA at 4°C.

Individual QDs blink and this well-known phenomenon enables one to distinguish single vs. multiple QDs (Dahan *et al.* 2003; Yao *et al.* 2005; Zhang *et al.* 2009). In our experiments a non-blinking QD signal could arise from either QD aggregation or protein aggregation, therefore any QDs that did not blink were discarded from analysis, ensuring all reported data arose from single fluorescent molecules. Apparent differences in signal intensities within the kymograms can arise from several sources, including: normal variations in QD fluorescence; variations in QDs blinking frequency; stationary QDs are brighter because their signal is confined to a fixed location during frame acquisition; nonblinking signals are brighter because they arise from multiple QDs; and some faint signals in the kymograms arise from “bleed-through” of signal from QD-proteins bound to an adjacent DNA in the curtain. These variations are expected, and we have confined our analysis and

calculations to blinking QDs to ensure we are monitoring single fluorescent molecules.

Two-color labeling to assess the oligomeric state of Mlh1-Pms1. Two-color QD labeling experiments (as described in (Reck-Peterson *et al.* 2006) were used to assess whether the Mlh1-Pms1 complexes under investigation were consistent with single heterodimers or higher order oligomers. In brief, we premixed equimolar amounts of anti-Flag green QDs (Qdot® 585) and anti-Flag magenta QDs (Qdot® 705), and then used this two-color mixture to label Mlh1(Flag)-Pms1. Heterodimeric Mlh1(Flag)-Pms1 contains just one Flag epitope, and therefore could only be labeled with one QD (either green OR magenta, but not both). Whereas higher order oligomers would contain multiple Flag tags (the exact number of Flag tags would scale in proportion to the number of Mlh1-Pms1 subunits within the oligomer) and therefore could be labeled with two or more QDs, leading to colocalization of both green and magenta QDs. These two-color colocalization experiments revealed that 94.8% (N=1,254/1,323) of the proteins were either only green or only magenta, arguing that most Mlh1-Pms1 complexes contained only a single Flag epitope. This result is consistent with a heterodimeric Mlh1-Pms1 complex, but inconsistent with the formation of larger Mlh1-Pms1 oligomers at the low protein concentrations used for our diffusion experiments.

Double-tethered DNA Curtains. A complete description of the double-tethered DNA curtains can be found in (Gorman *et al.* 2010). Fused silica slides (G. Finkenbeiner, Inc.) were cleaned in NanoStrip solution (CyanTek Corp, Fremont, CA) for 20 minutes, rinsed with acetone and isopropanol and dried with N_2. Slides were spin-coated with a bilayer of polymethylmethacrylate (PMMA; 25K and 495K; MicroChem, Newton, MA), followed by a

layer of Aquasave (Mitsubishi Rayon). Patterns were written with a FEI Sirion scanning electron microscope (J. C. Nabity, Inc., Bozeman, MT). Aquasave was removed with deionized water and resist was developed using isopropanol:methyl isobutyl ketone (3:1) for 1 minute with ultrasonic agitation at 5°C. The substrate was rinsed in isopropanol and dried with N_2. Barriers were made with a 15-20 nm layer of chromium (Cr), and following liftoff, samples were rinsed with acetone and dried with N_2, as described (Gorman *et al.* 2010).

Inlet and outlet ports were made by boring through the slide with a precision drill press equipped with a diamond-tipped bit (1.4 mm O.D.; Kassoy). The slides were cleaned by successive immersion in 2% (v/v) Hellmanex, 1 M NaOH, and 100% MeOH. Slides were rinsed with MilliQ™ between each wash and stored in 100% MeOH until use. Prior to assembly, slides were dried under a stream of nitrogen and baked in a vacuum oven for at least 1 hour. A sample chamber was prepared from a borosilicate glass coverslip (Fisher Scientific) and double-sided tape (~100 μm thick, 3M). Ports (Upchurch Scientific) were attached with hot-melt adhesive (SureBonder glue sticks, FPC Corp.). The total volume of the sample chambers was ~13 μl. A syringe pump (Kd Scientific) and actuated injection valves (Upchurch Scientific) were used to control sample delivery. The flowcell and prism were mounted in a custom-built heater with computer-controlled feedback regulation.

Lipids were purchased from Avanti Polar Lipids and liposomes were prepared as previously described (Gorman *et al.* 2010). In brief, a mixture of DOPC (1,2-dioleoyl-*sn*-glycerophosphocholine), 0.5% biotinylated-DPPE (1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine-N-(cap biotinyl)), and 8% mPEG 550-DOPE (1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-550]). The mPEG prevented nonspecific adsorption of QDs. Liposomes were applied to the sample chamber for 15

minutes. Excess liposomes were removed with buffer containing 10 mM Tris-HCl [pH 7.8] and 100 mM NaCl. The flowcell was then rinsed with the same buffer and incubated for 30 minutes. 30 μg/ml anti-DIG Fab (Roche Cat. No. 1214667001) or anti-FITC (Invitrogen Cat. No. 71-1900) was injected into the chamber and incubated for 20 minutes. The sample chamber was then flushed with buffer A (40 mM Tris-HCl [pH 7.8], 1 mM DTT, and 1 mM $MgCl_2$) plus 0.2 mg/ml BSA for 5 minutes. Streptavidin (0.02 mg/ml) in buffer A was injected into the sample chamber and incubated for 20 minutes. After rinsing with additional buffer A plus 0.2 mg/ml BSA, λ-DNA (15-20 pM) labeled at one end with biotin and at the other end with DIG or FITC and pre-stained with 0.5 nM YOYO1 was injected into the chamber, incubated for 10 minutes, and unbound DNA was removed by flushing with buffer at 0.1 ml/min. Application of flow aligned the DNA molecules along the diffusion barriers, and stretched the molecules so the free ends could attach to the pentagons.

Using hydrodynamic force to push proteins. All experiments where proteins were pushed along DNA by hydrodynamic force were done at a flow rate of 0.5 ml/min. When considering flow through a channel in which the height (h) is much less than the width (w), one can use the Navier-Stokes equations to determine the flow profile $v(y)$ obtained when a pressure difference is applied between the two ends of the channel. In this case the top and bottom surfaces of the flowcell chamber create a drag on the buffer that results in a parabolic flow where the velocity of the buffer at a distance y from the surface can be described by the equation:

$$v(y) = 4v_m/h^2\ (hy - y^2),$$

where v_m is the maximum velocity in the middle of the channel (Berg 1993). The maximum

velocity can be determined by reasoning that the volume of total fluid passing through the channel per unit of time equals:

$$w\int_0^h v(y)\partial y = 2/3whv_m.$$

With the velocity obtained from the above equations the force acting upon a molecule can then be determined by Stokes law which states that to move a molecule of radius *a* at a velocity of v_d in buffer of a viscosity η the force required is equal to:

$$F = 6\pi\eta avd.$$

We approximated the distance *y* from the surface to be 100 nm and the radius of the protein-QD complex to be 13 nm. The height of a typical channel was measured to be 100 μm and the width of the channel was measured to be 4,500 μm. The resulting calculated force applied to a protein-QD complex at a buffer flow of 0.5 ml/min is approximately 25 fN. At 200 nm from the surface this theoretical value increases to approximately 50 fN.

We then determined the force acting on a complex experimentally and compared these experimental results to the above theoretical calculations. The drift velocity of a particle is equal to the force exerted on that molecule (*F*) divided by the drag coefficient of the molecule (*f*)(Berg 1993):

$$v_d = F/f$$

The diffusion coefficient (*D*) of a molecule is also determined by this drag coefficient:

$$D = kT/f.$$

We were therefore able to calculate the force applied to a single Mlh1-QD complex on DNA by first determining its diffusion coefficient and subsequently pushing the molecule with buffer flow and tracking the movement to determine its drift velocity. Using these values we determined the actual force on the molecule to be 100 fN, which is in close agreement to the

theoretical values calculated above. Importantly, this experimental method implicitly includes the drag components from the hydrodynamic radius of the complex as well as the protein DNA interaction, without needing to determine these components directly, and also makes no assumptions regarding the distance of the protein from the surface.

Experiments with chromatin substrates. Nucleosomes were prepared as described (Visnapuu and Greene 2009). Histones (H2A, Flag-H2B, H3, and H4) were expressed in *E. coli*, purified from inclusion bodies and reconstituted as described (Wittmeyer *et al.* 2004). In brief, inclusion bodies were resuspended in unfolding buffer (7 M guanidinium-HCl, 1 M NaCl, 50 mM Tris-HCl [pH 7.8], 1 mM EDTA, 1 mM DTT), dialyzed against urea buffer (7 M urea, 1 M NaCl, 10 mM Tris-HCl [pH 7.8], 1 mM EDTA, 5 mM β-mercaptoethanol), then loaded onto tandem HiTrap Q and SP columns (GE Healthcare). Histones were eluted from the SP column with a 100-400 mM NaCl gradient for H2A and Flag-H2B and a 200-500 mM NaCl gradient for H3 and H4. Purified histones were dialysed against 10 mM Tris-HCl [pH 7.8] plus 5 mM β-mercaptoethanol, followed by 10 mM Tris-HCl [pH 7.8], then lyophilized and stored at –20°C. Lyophilized histones were unfolded in 7M guanidinium-HCl, 1 M NaCl, 50 mM Tris-HCl [pH 7.8] plus 10 mM DTT, combined at equimolar ratios, and dialyzed into 2 M NaCl, 20 mM Tris-HCl [pH 7.8], 1 mM EDTA, 5 mM β-mercaptoethanol with several buffer changes over 48 hours. Reconstituted octamers were purified by gel filtration and deposited onto DNA by salt dialysis (Luger *et al.* 1999; Thastrom *et al.* 2004).

The Flag-tagged nucleosomes were labeled with 0.5 nM QDs (QD 585, Invitrogen) conjugated to anti-Flag antibodies (Sigma). Nucleosome labeling was done *in situ* as described (Visnapuu and Greene 2009), either before or after the injection of Mlh1-Pms1, as

indicated (see Figure 2.4). For experiments with unlabeled nucleosomes, we first conducted the diffusion experiment using QD-Mlh1-Pms1 (QD 705, Invitrogen), in buffer containing 40 mM Tris-HCl [pH 7.7], 150 mM NaCl, ±1 mM ATP, 0 mM Mg^{2+}, 1 mM DTT, and 0.4 mg/ml BSA. Mg^{2+} induces condensation of the chromatin substrates by promoting nucleosome-nucleosome interactions (M.-L.V. and J.G., unpublished), and for this reason was omitted from the diffusion experiments. Mlh1-Pms1 was then flushed from the chamber with 300 mM NaCl (which does not disrupt the nucleosomes; M-L.V and E.C.G, unpublished, (Burton *et al.* 1978; Park *et al.* 2004) and the nucleosomes were then located by labeling with QDs (QD 705, Invitrogen). The QD signal from the labeled nucleosomes was then pseudocolored and superimposed on the kymograms of Mlh1-Pms1 diffusion (see Figure 2.4A, B, upper and middle panels). For reactions with labeled nucleosomes, the nucleosomes were labeled *in situ* with anti-Flag QDs (QD 585, Invitrogen), then QD-Mlh1-Pms1 (QD 705, Invitrogen) was injected into the sample chamber and the signal from the different colored QDs was collected concurrently; signal gaps in the real time data correspond to QD blinking (see Figure 2.4A, B, lower panels).

The chromatin diffusion experiments with Msh2-Msh6 were conducted essentially the same as with Mlh1-Pms1, in buffer containing 40 mM Tris-HCl [pH 7.7], 50 mM NaCl, 1 mM ADP, 1 mM DTT, and 0.4 mg/ml BSA. ATP was omitted to prevent ATP-triggered protein dissociation, and the salt concentration was kept at 50 mM NaCl to increase the half-life of the bound state in order to evaluate whether it was able to pass nucleosomes (Gorman *et al.* 2007). Under identical reaction conditions, in the absence of nucleosomes, many molecules (48%, N=380) of Msh2-Msh6 reversibly enter a nondiffusive state (immobile) thought to mimic a pseudo-damage recognition complex, and the remaining Msh2-Msh6

complexes remain mobile (52%, N=412; (Gorman *et al.* 2007). In the presence of high-density nucleosomes (~80-100 nucleosomes per DNA molecule), the fraction of immobile molecules increases from 48% to 96.4%, and we attribute this increase to Msh2-Msh6 molecules that are now trapped between nucleosomes and incapable of diffusing on DNA.

Data analysis. All diffusion coefficients represent the mean ± standard deviation of ≥25 particle tracking measurements and were calculated from MSD plots as described in (Gorman *et al.* 2007). Traces were used to calculate diffusion coefficients only if the QDs could be tracked over ≥250 consecutive frames (50 seconds), only QDs that blinked were used for tracking (verifying they were single QDs), and traces were excluded if collisions between two or more proteins prevented accurate tracking. We can calculate diffusion coefficients using fewer than 250 consecutive frames, but the variance and error in the resulting data begins to increase significantly (not shown). Trajectories where the proteins approached to within 500-nm of one another were also excluded from the diffusion coefficient calculations as a quality control measure because error in the tracking algorithm increases below this distance. Two closely approaching QDs can be optically resolved from one another at distances >10 nm (Lacoste *et al.* 2000; Lagerholm *et al.* 2006), but as indicated above, as a quality control measure we excluded any traces where proteins approached to within 500 nm of one another to ensure uniform accuracy in the particle tracking data.

Diffusion coefficients were calculated from the tracking data as previously described (Gorman *et al.* 2007). In brief, the movement of each protein complex in the y-direction (parallel to the long axis of the DNA) was then analyzed to calculate the mean squared displacement (MSD) using:

$$MSD(n\Delta T) = \sum_{i=0}^{N} (Y_{i+n} - Y_i)^2 / (N+1)$$

where $n\Delta T$ = 10% of the total diffusion time (to minimize errors due to sampling size) (Qian *et al.* 1991). Using the MSD information, the diffusion coefficient for each protein complex was calculated by:

$$D(t) = MSD(t) / 2t$$

where $D(t)$ is the diffusion coefficient for time interval t (Qian *et al.* 1991); Berg, 1993). For the linear MSD traces, the diffusion coefficients were calculated from direct fits to the entire plot. For nonlinear MSD plots the diffusion coefficients were estimated from the initial slope of the curve and this slope is not appreciably affected by nondiffusive behavior or bounded diffusion (Kusumi *et al.* 1993; Saxton and Jacobson 1997). The 1D-diffusion coefficients display a lognormal distribution, which likely arises due to the roughness of the energy landscape (Gorman *et al.* 2007), and where indicated student t-tests were performed on the natural logarithm (ln) of the diffusion coefficients to obtain p-values for statistical comparisons of the data.

The particle-tracking algorithm used to monitor the movement of Msh2-Msh6 and Mlh1-Pms1simultaneously records its position in the y-direction (parallel to the long axis of the DNA) and in the x-direction (perpendicular to the long axis of the DNA). The values obtained for the x-direction primarily reflect the thermal motions of the DNA molecules. All of the tracked proteins displayed x-direction fluctuations ranging between ±50-250 nanometers (with a mean of ±80 nm), which is an order of magnitude below the motions observed along the helical axis (i.e. the y-direction) of the DNA. DNAs more flexible than this were occasionally observed, but they were omitted from the analysis because their

flexibility caused the bound fluorescent Msh2-Msh6 to fluctuate too much within the evanescent field making them impossible to track accurately (although qualitatively they displayed exactly the same diffusive behavior). The transverse fluctuations of the DNA and the temporal resolution of our detection system impose a lower limit of $>1x10^{-4}$ μm^2/sec for the diffusion coefficients that can be measured (anything slower than this will look like a stationary particle), but this lower limit is well below any values reported in our study.

Acknowledgements

We thank members of the Alani and Greene laboratories for carefully reading the manuscript and providing suggestions throughout this study. This work was supported by the Howard Hughes Medical Institute, by a US National Science Foundation Presidential Early Career Award for Scientists and Engineers, by US National Institutes of Health (NIH) grant GM082848 to E.C.G. and by NIH grant GM53085 to E.A. J.G. is supported by an NIH training grant for Cellular and Molecular Foundations of Biomedical Sciences (T32GM00879807). A.J.P. was supported by a State University of New York fellowship. This work was partially supported by the Initiatives in Science and Engineering program through Columbia University, the Nanoscale Science and Engineering Initiative of the National Science Foundation under US National Science Foundation Award Number CHE-0641523 and by the New York State Office of Science, Technology, and Academic Research.

References

Bagchi, B., P. C. Blainey, et al. (2008). "Diffusion constant of a nonspecifically bound protein undergoing curvilinear motion along DNA." J Phys Chem B **112**(19): 6282-4.

Berg, H. C. (1993). Random walks in biology. Princeton, Princeton University Press.

Berg, O. G., R. B. Winter, et al. (1981). "Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory." Biochemistry **20**(24): 6929-48.

Blainey, P. C., A. M. van Oijen, et al. (2006). "A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA." Proc Natl Acad Sci U S A **103**: 5752-57.

Burton, D. R., M. J. Butler, et al. (1978). "The interaction of core histones with DNA: equilibrium binding studies." Nucleic Acids Res **5**(10): 3643-63.

Chi, N. W. and R. D. Kolodner (1994). "Purification and characterization of MSH1, a yeast mitochondrial protein that binds to DNA mismatches." J Biol Chem **269**(47): 29984-92.

Dahan, M., S. Levi, et al. (2003). "Diffusion dynamics of glycine receptors revealed by single-quantum dot tracking." Science **302**(5644): 442-5.

Doucleff, M. and G. M. Clore (2008). "Global jumping and domain-specific intersegment transfer between DNA cognate sites of the multidomain transcription factor Oct-1." Proc Natl Acad Sci U S A **105**(37): 13871-6.

Elf, J., G. W. Li, et al. (2007). "Probing transcription factor dynamics at the single-molecule level in a living cell." Science **316**(5828): 1191-4.

Gorman, J., A. Chowdhury, et al. (2007). "Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6." Mol Cell **28**(3): 359-70.

Gorman, J., T. Fazio, et al. (2010). "Nanofabricated racks of aligned and anchored DNA substrates for single-molecule imaging." Langmuir **26**(2): 1372-9.

Gorman, J. and E. C. Greene (2008). "Visualizing one-dimensional diffusion of proteins along DNA." Nat Struct Mol Biol **15**(8): 768-74.

Gorski, S. A., M. Dundr, et al. (2006). "The road much traveled: trafficking in the cell nucleus." Curr Opin Cell Biol **18**(3): 284-90.

Gradia, S., D. Subramanian, et al. (1999). "hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA." Mol Cell **3**(2): 255-61.

Groth, A., W. Rocha, et al. (2007). "Chromatin challenges during DNA replication and repair." Cell **128**(4): 721-33.

Guarne, A., M. S. Junop, et al. (2001). "Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase." Embo J **20**(19): 5521-31.

Guarne, A., S. Ramon-Maiques, et al. (2004). "Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair." Embo J **23**(21): 4134-45.

Hager, G. L., J. G. McNally, et al. (2009). "Transcription dynamics." Mol Cell **35**(6): 741-53.

Halford, S. E. and M. D. Szczelkun (2002). "How to get from A to B: strategies for analysing protein motion on DNA." Eur Biophys J **31**(4): 257-67.

Hall, M. C. and T. A. Kunkel (2001). "Purification of eukaryotic MutL homologs from

Saccharomyces cerevisiae using self-affinity technology." Protein Expr Purif **21**(2): 333-42.

Hall, M. C., P. V. Shcherbakova, et al. (2003). "DNA binding by yeast Mlh1 and Pms1: implications for DNA mismatch repair." Nucleic Acids Res **31**(8): 2025-34.

Hall, M. C., H. Wang, et al. (2001). "High affinity cooperative DNA binding by the yeast Mlh1-Pms1 heterodimer." J Mol Biol **312**(4): 637-47.

Heck, J. A., J. L. Argueso, et al. (2006). "Negative epistasis between natural variants of the Saccharomyces cerevisiae MLH1 and PMS1 genes results in a defect in mismatch repair." Proc Natl Acad Sci U S A **103**(9): 3256-61.

Hodges, C., L. Bintu, et al. (2009). "Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II." Science **325**(5940): 626-8.

Iwahara, J. and G. M. Clore (2006). "Direct observation of enhanced translocation of a homeodomain between DNA cognate sites by NMR exchange spectroscopy." J Am Chem Soc **128**(2): 404-5.

Iwahara, J., M. Zweckstetter, et al. (2006). "NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA." Proc Natl Acad Sci U S A **103**(41): 15062-7.

Jiricny, J. (2006). "The multifaceted mismatch-repair system." Nat Rev Mol Cell Biol **7**(5): 335-46.

Kampmann, M. (2005). "Facilitated diffusion in chromatin lattices: mechanistic diversity and regulatory potential." Mol Microbiol **57**(4): 889-99.

Kijas, A. W., B. Studamire, et al. (2003). "Msh2 separation of function mutations confer defects in the initiation steps of mismatch repair." J Mol Biol **331**(1): 123-38.

Kolodner, R. D., M. L. Mendillo, et al. (2007). "Coupling distant sites in DNA during DNA mismatch repair." Proc Natl Acad Sci U S A **104**(32): 12953-4.

Komazin-Meredith, G., R. Mirchev, et al. (2008). "Hopping of a processivity factor on DNA revealed by single-molecule assays of diffusion." Proc Natl Acad Sci U S A **105**(31): 10721-6.

Kosinski, J., I. Steindorf, et al. (2005). "Analysis of the quaternary structure of the MutL C-terminal domain." J Mol Biol **351**(4): 895-909.

Kunkel, T. A. and D. A. Erie (2005). "DNA mismatch repair." Annu Rev Biochem **74**: 681-710.

Kusumi, A., Y. Sako, et al. (1993). "Confined lateral diffusion of membrane receptors as studied by single particle tracking (nanovid microscopy). Effects of calcium-induced differentiation in cultured epithelial cells." Biophys J **65**(5): 2021-40.

Lacoste, T. D., X. Michalet, et al. (2000). "Ultrahigh-resolution multicolor colocalization of single fluorescent probes." Proc Natl Acad Sci U S A **97**(17): 9461-6.

Lagerholm, B. C., L. Averett, et al. (2006). "Analysis method for measuring submicroscopic distances with blinking quantum dots." Biophys J **91**(8): 3050-60.

Li, F., L. Tian, et al. (2009). "Evidence that nucleosomes inhibit mismatch repair in eukaryotic cells." J Biol Chem **284**(48): 33056-61.

Li, G.-W., O. Berg, et al. (2009). "Effects of macromolecular crowding and DNA looping on

gene regulation kinetics." Nature Physics **5**: 294-297.

Liu, S., E. A. Abbondanzieri, et al. (2008). "Slide into action: dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates." Science **322**(5904): 1092-7.

Luger, K., A. W. Mader, et al. (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-60.

Luger, K., T. J. Rechsteiner, et al. (1999). "Preparation of nucleosome core particle from recombinant histones." Methods Enzymol **304**: 3-19.

Mendillo, M. L., D. J. Mazur, et al. (2005). "Analysis of the interaction between the Saccharomyces cerevisiae MSH2-MSH6 and MLH1-PMS1 complexes with DNA using a reversible DNA end-blocking system." J Biol Chem **280**(23): 22245-57.

Mirny, L. A., M. Slutsky, et al. (2009). "How a protein searches for its site on DNA: the mechanism of facilitated diffusion." J Phys A: Math Theor. **42**(43): 1-23.

Modrich, P. (2006). "Mechanisms in eukaryotic mismatch repair." J Biol Chem **281**(41): 30305-9.

Park, Y. J., P. N. Dyer, et al. (2004). "A new fluorescence resonance energy transfer approach demonstrates that the histone variant H2AZ stabilizes the histone octamer within the nucleosome." J Biol Chem **279**(23): 24274-82.

Pathak, S., M. C. Davidson, et al. (2007). "Characterization of the functional binding properties of antibody conjugated quantum dots." Nano Lett **7**(7): 1839-45.

Pluciennik, A. and P. Modrich (2007). "Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair." Proc Natl Acad Sci U S A **104**(31):

12709-13.

Qian, H., M. P. Sheetz, et al. (1991). "Single particle tracking. Analysis of diffusion and flow in two-dimensional systems." Biophys J **60**(4): 910-21.

Reck-Peterson, S. L., A. Yildiz, et al. (2006). "Single-molecule analysis of dynein processivity and stepping behavior." Cell **126**(2): 335-48.

Riggs, A. D., S. Bourgeois, et al. (1970). "The lac repressor-operator interaction. 3. Kinetic studies." J Mol Biol **53**(3): 401-17.

Roy, R., A. G. Kozlov, et al. (2009). "SSB protein diffusion on single-stranded DNA stimulates RecA filament formation." Nature **461**(7267): 1092-7.

Sacho, E. J., F. A. Kadyrov, et al. (2008). "Direct visualization of asymmetric adenine-nucleotide-induced conformational changes in MutL alpha." Mol Cell **29**(1): 112-21.

Saxton, M. J. and K. Jacobson (1997). "Single-particle tracking: applications to membrane dynamics." Annu Rev Biophys Biomol Struct **26**: 373-99.

Studitsky, V. M., D. J. Clark, et al. (1995). "Overcoming a nucleosomal barrier to transcription." Cell **83**(1): 19-27.

Surtees, J. A. and E. Alani (2006). "Mismatch repair factor MSH2-MSH3 binds and alters the conformation of branched DNA structures predicted to form during genetic recombination." J Mol Biol **360**(3): 523-36.

Tafvizi, A., F. Huang, et al. (2008). "Tumor suppressor p53 slides on DNA with low friction and high stability." Biophys J **95**(1): L01-3.

Thastrom, A., P. T. Lowary, et al. (2004). "Measurement of histone-DNA interaction free energy in nucleosomes." Methods **33**(1): 33-44.

Umar, A., A. B. Buermeyer, et al. (1996). "Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis." Cell **87**(1): 65-73.

Visnapuu, M. L. and E. C. Greene (2009). "Single-molecule imaging of DNA curtains reveals intrinsic energy landscapes for nucleosome deposition." Nat Struct Mol Biol **16**(10): 1056-62.

von Hippel, P. H. and O. G. Berg (1989). "Facilitated target location in biological systems." J Biol Chem **264**(2): 675-8.

Warren, J. J., T. J. Pohlhaus, et al. (2007). "Structure of the human MutSalpha DNA lesion recognition complex." Mol Cell **26**(4): 579-92.

Winter, R. B., O. G. Berg, et al. (1981). "Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor--operator interaction: kinetic measurements and conclusions." Biochemistry **20**(24): 6961-77.

Winter, R. B. and P. H. von Hippel (1981). "Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The Escherichia coli repressor--operator interaction: equilibrium measurements." Biochemistry **20**(24): 6948-60.

Wittmeyer, J., A. Saha, et al. (2004). "DNA translocation and nucleosome remodeling assays by the RSC chromatin remodeling complex." Methods Enzymol **377**: 322-43.

Yao, J., D. R. Larson, et al. (2005). "Blinking and nonradiant dark fraction of water-soluble quantum dots in aqueous solution." Proc Natl Acad Sci U S A **102**(40): 14284-9.

Zhang, Q., Y. Li, et al. (2009). "The dynamic control of kiss-and-run and vesicular reuse probed with single nanoparticles." Science **323**(5920): 1448-53.

Chapter 3

The function of unstructured linker arm domains of MLH1 and PMS1 is consistent with enabling diffusion of the complex along chromatin

Introduction

DNA binding proteins have the challenging task of locating specific targets among a vast excess of non-specific DNA. Nonetheless, proteins such as Lac repressor bind their targets at rates exceeding theoretical predictions for random collisions in three-dimensional space (Riggs *et al.* 1970; Berg *et al.* 1981; Winter *et al.* 1981; Winter and von Hippel 1981). The ability of proteins to quickly and efficiently find their targets has been described as facilitated diffusion; the area being searched is limited to the one-dimensional trajectory along DNA (von Hippel and Berg 1989). Models for facilitated diffusion purport several mechanisms for protein movement along DNA that would explain the rapid rates observed including: sliding, hopping, intersegmental transfer and jumping (Halford and Marko 2004). However, one very important aspect of diffusion along DNA is how proteins deal with obstacles such as nucleosomes. In the absence of mechanical means to disrupt nucleosomes, diffusing proteins must be able to circumvent these potential barriers through other mechanisms. Recently, the DNA mismatch repair complex MLH1-PMS1 was observed to freely bypass nucleosomes while employing a hopping/stepping one-dimensional diffusion mechanism (Gorman *et al.* 2010). In contrast, diffusion of the DNA mismatch recognition complex MSH2-MSH6, which utilizes a sliding mechanism (Gorman *et al.* 2007), was blocked by stably bound nucleosomes. The nature of the difference between the two

complexes in dealing with nucleosome barriers is likely due to how they interact with the DNA as they move (i.e. sliding while tracking the phosphate backbone vs. making microscopic association/disassociations).

Structural studies on members of the MLH family have revealed important insights into how these proteins might bind and move along DNA (Ban and Yang 1998; Ban *et al.* 1999; Guarne *et al.* 2001; Guarne *et al.* 2004; Kosinski *et al.* 2005). N-terminal domains (NTDs) of MutL family members are highly conserved and contain an ATP binding site that belongs to the GHKL family of ATPases (Dutta and Inouye 2000; Sacho *et al.* 2008). Upon nucleotide binding, NTDs facilitate large conformational changes in the linker arms that connect to the C-terminal domains (CTDs) (Sacho *et al.* 2008). The less conserved CTDs are thought to be structurally similar and are essential for dimerization (Hall *et al.* 2003). The linker arm that connects the two terminal globular domains has been found to be devoid of secondary structure, variable in length between MutL family members and highly divergent in sequence context (Guarne *et al.* 2004; Sacho *et al.* 2008). Failed attempts at crystallizing full-length MutL family members along with secondary structure prediction programs suggest that the linker arms are random coils that are highly disordered in solution (Guarne *et al*. 2004). Single-molecule analysis in combination with structural information strongly suggests that *Saccharomyces cerevisiae* MLH1-PMS1 complexes adopt a ring-like configuration capable of encircling DNA while using a rapid hopping/stepping diffusion mechanism (Guarne *et al.* 2004; Gorman *et al.* 2010). This mechanism is thought to involve the long linker arms of the two subunits (~160 amino acids for MLH1 and ~300 amino acids for PMS1) coordinating successive hops/steps along DNA and over obstacles to mismatch repair (i.e. nucleosomes).

DNA mismatch repair is a conserved pathway that corrects replication errors

introduced by DNA polymerase (Reviewed in (Kunkel and Erie 2005). Utilization of this pathway improves the fidelity of DNA replication by several orders of magnitude (Modrich and Lahue 1996). Furthermore, mutations in the *MSH* and *MLH* mismatch repair genes results in significant increases in mutation rates and are frequently found in patients with hereditary non-polyposis colorectal cancer (Lynch *et al.* 2009). Mismatch repair in eukaryotic organisms starts with the formation of a ternary complex consisting of MSH2-MSH6 and MLH1-PMS1 at a mismatch. This results in the initiation of downstream repair events that act to excise the mismatch in a manner that maintains template strand information (Habraken *et al.* 1998; Modrich 2006); Chapter 1). Interactions with the replication machinery (e.g. PCNA) are hypothesized to recruit and enrich the localization of the mismatch repair complexes to progressing replication forks (Clark *et al.* 2000; Flores-Rozas *et al.* 2000; Lau and Kolodner 2003). As a consequence of enrichment at the replication fork, MSH2-MSH6 is thought to rapidly scan immediately behind the fork using a sliding diffusion mechanism to locate polymerase errors (Figure 3.1A). Importantly in this model the timeframe for MSH2-MSH6 target search coincides with transient nucleosome disruption by the passing replication machinery (Groth *et al.* 2007).

Nucleosome deposition onto nascent DNA occurs through direct transfer of nucleosomes from in front of the fork and de novo formation from newly synthesized histones (Sogo *et al.* 1986; Shibahara and Stillman 1999). Psoralen crosslinking experiments revealed that nucleosome deposition occurs up to 700 base pairs away from the replication fork (Sogo *et al.* 1986). Therefore, nucleosomes, which have been shown to be a barrier to MSH2-MSH6 diffusion *in vitro*, would not present an obstacle to mismatch recognition (Li *et al.* 2009; Gorman *et al.* 2010). Although several hundred base pairs may initially be clear of

nucleosomes, *in vitro* nucleosome assembly assays have shown that DNA compaction occurs in only a few seconds (Ladoux *et al.* 2000; Wagner *et al.* 2005). MLH1-PMS1 likely scans for MSH2-MSH6 bound to a DNA mismatch. Such a scan is likely to take place in an environment that would include redeposited nucleosomes (Figure 3.1B). The ability of MLH1-PMS1 complexes to bypass nucleosomes could thus allow for efficient search and repair (Figure 3.1C). The nature of the MSH2-MSH6, MLH1-PMS1 complex after mismatch binding that leads to downstream signaling has not been elucidated, although one-dimensional movement along DNA is thought to be utilized. Studies have shown that MSH2-MSH6 can undergo a nucleotide-dependent transformation into a hydrolysis-independent sliding clamp that disengages from the mismatch site, but these studies do not include the behavior of MLH1-PMS1 (Gradia *et al.* 1997; Gradia *et al.* 1999).

To gain a clearer understanding of the mechanism by which MLH1-PMS1 encircles and rapidly diffuses along DNA, I have created a series of deletions within the linker arm domains of both MLH1 and PMS1. I show that the linker arm of MLH1 is more sensitive to deletion than PMS1 by a genetic assay that measures mismatch repair function. Similarly, proteolytic cleavage of the linker arm of MLH1 leads to a loss of mismatch repair activity *in vivo* and I have previously shown that cleavage results in a loss of DNA binding activity *in vitro*. Purified complexes containing deletions in MLH1-PMS1 linker arms were analyzed and found to have differential defects in DNA binding and in the ability to form a ternary complex with MSH2-MSH6 at a mismatch. Single-molecule studies are necessary in order to determine if mismatch repair defective complexes that can still bind DNA display altered diffusion rates or an inability to bypass nucleosomes. These experiments are described in the Discussion.

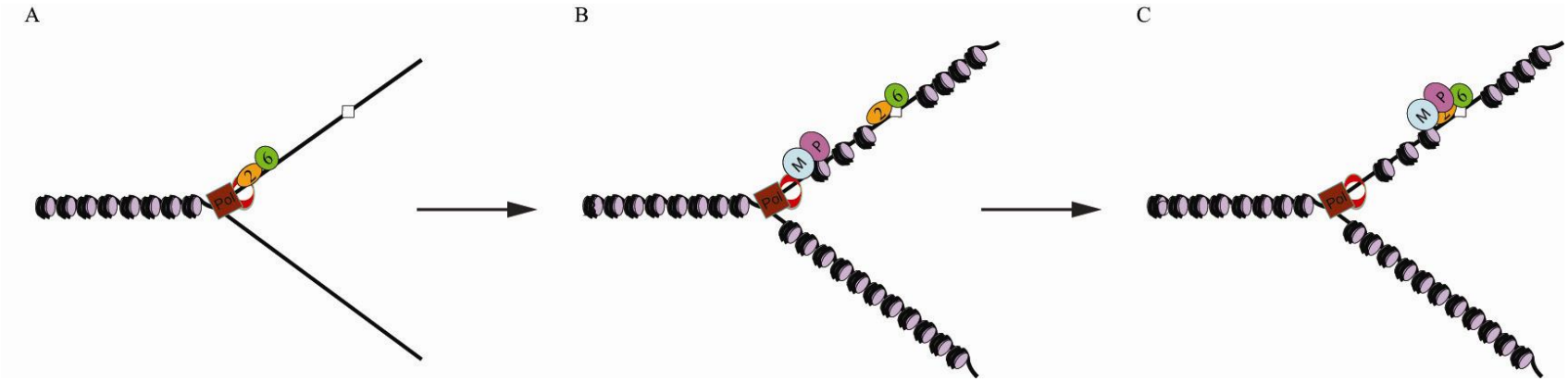


Figure 3.1. Model for replication-associated mismatch search by MSH2-MSH6 and MLH1-PMS1.
Model replication forks illustrating the environment for diffusion of each mismatch repair complex. **(A)** MSH2-MSH6 is potentially enriched at replication forks through direct physical association with the replication clamp PCNA (red disc). Nucleosomes downstream of the passing replication fork are transiently disrupted allowing MSH2-MSH6 to efficiently scan for newly created mismatches (diamond) without impediment. **(B)** Nucleosomes are quickly re-deposited downstream of the replication fork and stand as potential roadblocks to 1D protein diffusion along DNA. **(C)** Recently, it was shown that MLH1-PMS1 is capable of efficiently traversing nucleosomes and would be able to overcome these obstacles in order to form a ternary complex with MSH2-MSH6 at a mismatch (Gorman *et al.* 2010).

Results

Impairment of MLH1-PMS1 DNA binding through internal protease cleavage results in a mismatch repair defect *in vivo*. Using total internal reflection fluorescent microscopy (TIRFM) and *in vitro* TEV proteolytic cleavage assays, I showed that the intact linker arms of MLH1-PMS1 are required for stable DNA binding (Figure 3.2) (Gorman *et al.* 2010). These results and others in Chapter 2 (Figure 2.4) are consistent with MLH1-PMS1 adopting a ring-like configuration that wraps around DNA. To determine whether topological binding of DNA by MLH1-PMS1 is required for mismatch repair *in vivo*, I integrated *MLH1* alleles containing TEV cleavage sites into a strain background containing *TEV* protease with a nuclear localization signal under the galactose inducible promoter (Table 3.1) (Uhlmann *et al.* 2000). It should be noted that there are no proteins that contain the canonical TEV cleavage site in *Saccharomyces cerevisiae* and *TEV* expression does not have any discernable effects on growth and proliferation (Uhlmann *et al.* 2000; Kohler 2003). These strains also contain the *lys2*::*insE-A*$_{14}$ frameshift allele that serves as a measure of mismatch repair efficiency (Tran *et al.* 1997). Using this Lys$^+$ reversion assay, I tested the effects of TEV cleavage at two different sites (after amino acid 448T or 499Y) in the MLH1 linker arm. Strains bearing these two alleles displayed high mutation rates comparable to *mlh1Δ* in the presence of galactose, but were otherwise wild-type in the presence of the non-inducible carbon source sucrose (Table 3.2). In strains lacking TEV protease, these alleles fully complemented the *mlh1Δ* mutator phenotype. Elevation of Lys$^+$ reversion rates to the level observed for *mlh1Δ* were only obtained when cells were continuously maintained on media containing galactose as opposed to only when deposited on auxotrophy selection plates. I reasoned that this was

Figure 3.2. TEV cleavage of Mlh1-Pms1 linker arms disrupts DNA-binding.
(A) Schematic overview of different TEV-containing Mlh1-Pms1 constructs (Guarne *et al*. 2004, Kosinski *et al*. 2005). MLH1 is in magenta and PMS1 is in blue. Linker arms are illustrated by a series of unconnected dots. Relative position of the TEV cleavage site (black dashed line), FLAG-tag (Black circle) and HA-tag (Black star) are shown. For details on exact position of each tag see Material and Methods. **(B)** Coomassie-stained SDS-PAGE showing specificity of TEV cleavage for each of the different constructs. **(C)** Gel shift assays using a ^{32}P-labeled oligonucleotide substrate ±TEV cleavage. All of the proteins bind DNA before TEV cleavage, but treatment with TEV protease reduces or eliminates DNA binding activity in the bulk assay. Similarly, all of these protein constructs bound and diffused on DNA in TIRFM assays, but no DNA binding activity was detected in the TIRFM assays after TEV cleavage of the linker arms (data not shown).

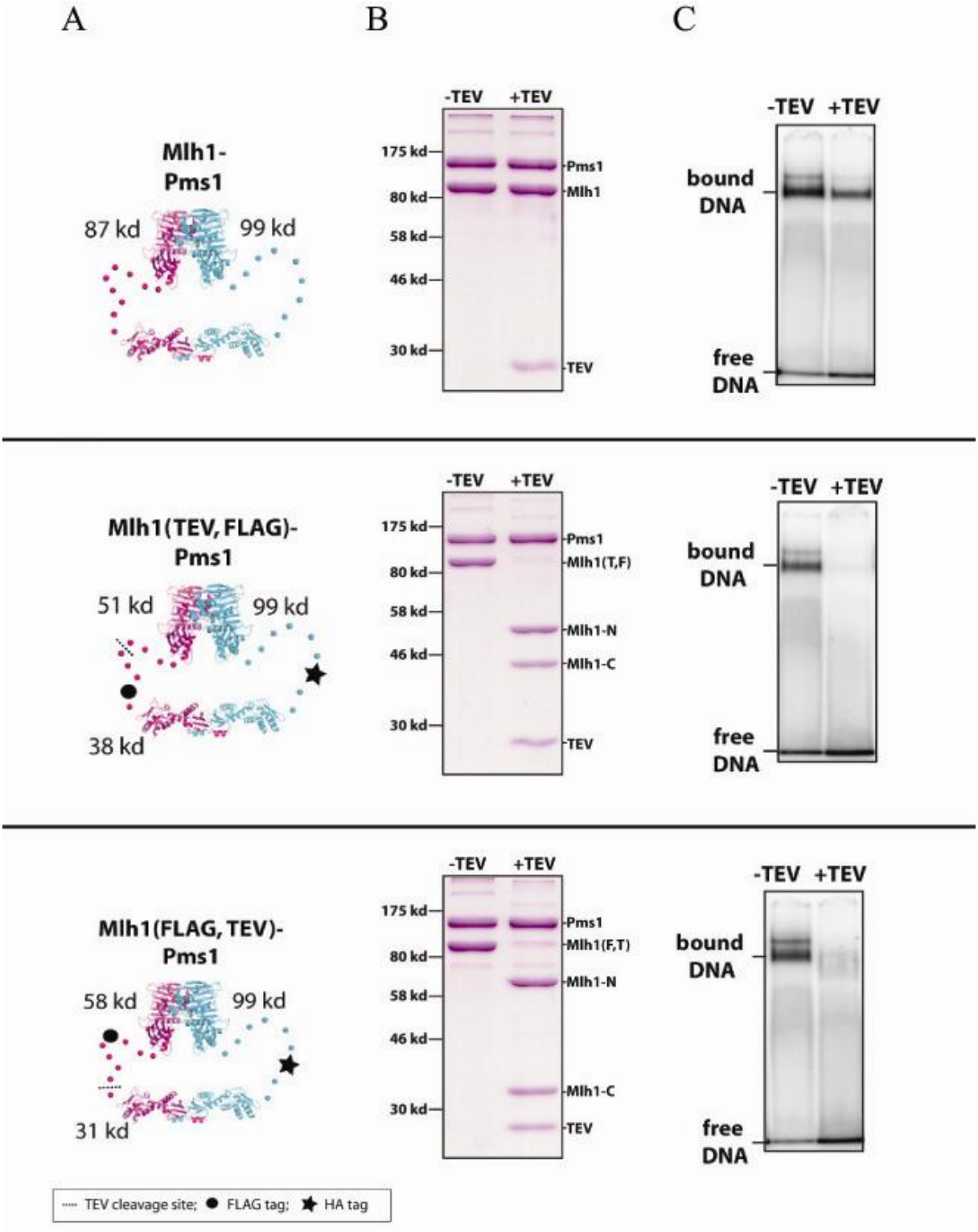

A
B
C
Mlh1-
Pms1
87 kd
99 kd
-TEV
+TEV
175 kd
80 kd
58 kd
46 kd
30 kd
Pms1
Mlh1
TEV
-TEV
+TEV
bound
DNA
free
DNA
Mlh1(TEV, FLAG)-
Pms1
51 kd
99 kd
38 kd
-TEV
+TEV
175 kd
80 kd
58 kd
46 kd
30 kd
Pms1
Mlh1(T,F)
Mlh1-N
Mlh1-C
TEV
-TEV
+TEV
bound
DNA
free
DNA
Mlh1(FLAG, TEV)-
Pms1
58 kd
99 kd
31 kd
-TEV
+TEV
175 kd
80 kd
58 kd
46 kd
30 kd
Pms1
Mlh1(F,T)
Mlh1-N
Mlh1-C
TEV
-TEV
+TEV
bound
DNA
free
DNA
TEV cleavage site; FLAG tag; HA tag

Table 3.1

List of strains and plasmids used in this study

Strains

EAY907 *MATa, ura3-52, leu2-3, 112, trp1-289, prb1-1122, prc1-407, pep4-3*

EAY1365 *MATa, ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2::insE-A14, mlh1Δ::KanMX4, pms1Δ::KanMX4*

EAY1366 *MATa, ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2::insE-A14, mlh1Δ::KanMX4*

EAY2576 *MATa, ura3, leu2-3, 112, omns GAL-NLS-myc9-TEV protease-NLS2::TRP1(10-fold integrant by southern)*

EAY3097 *MATa, ura3, leu2, trp1, his3Δ200, lys2::insE-A14, pms1Δ::KanMX4*

EAY3098 *MATa, ura3, leu2-3, 112, omns GAL-NLS-myc9-TEV protease-NLS2::TRP1(10-fold integrant by southern), mlh1Δ::KanMX4*

EAY3099 *MATa, ura3, leu2-3, 112, omns GAL-NLS-myc9-TEV protease-NLS2::TRP1(10-fold integrant by southern), MLH1::KanMX4*

EAY3100 *MATa, ura3, leu2-3, 112, omns GAL-NLS-myc9-TEV protease-NLS2::TRP1(10-fold integrant by southern), flag-MLH1::KanMX4*

EAY3101 *MATa, ura3, leu2-3, 112, omns GAL-NLS-myc9-TEV protease-NLS2::TRP1(10-fold integrant by southern), flag-tev-MLH1::KanMX4*

EAY3102 *MATa, ura3, leu2-3, 112, omns GAL-NLS-myc9-TEV protease-NLS2::TRP1(10-fold integrant by southern), tev-flag-MLH1::KanMX4*

Plasmids	**Relevant genotype**	**Vector type**
pEAA213	*MLH1::KanMX4*	*ARS-CEN*, integrating
pEAA238	*PMS1*	*ARS-CEN*
pEAA373	*flag-MLH1*	*ARS-CEN*, integrating
pEAA375	*flag-MLH1*	*ARS-CEN*, integrating
pEAA515	*flag-tev-MLH1*	*ARS-CEN*, integrating
pEAA516	*tev-flag-MLH1*	*ARS-CEN*, integrating
pEAA517	*ha-PMS1*	*ARS-CEN*
pEAA526	*mlh1Δ348-373*	*ARS-CEN*, integrating
pEAA527	*mlh1Δ445-470*	*ARS-CEN*, integrating
pEAA528	*mlh1Δ359-409*	*ARS-CEN*, integrating
pEAA529	*mlh1Δ407-457*	*ARS-CEN*, integrating
pEAA530	*mlh1Δ357-457*	*ARS-CEN*, integrating

pEAA531	*mlh1Δ336-480*	*ARS-CEN*, integrating
pEAA532	*mlh1Δ396-421*	*ARS-CEN*, integrating
pEAA544	*pms1Δ450-475*	*ARS-CEN*
pEAA545	*pms1Δ600-625*	*ARS-CEN*
pEAA546	*pms1Δ437-487*	*ARS-CEN*
pEAA547	*pms1Δ511-561*	*ARS-CEN*
pEAA548	*pms1Δ584-634*	*ARS-CEN*
pEAA549	*pms1Δ450-550*	*ARS-CEN*
pEAA550	*pms1Δ390-610*	*ARS-CEN*
pEAE269	*GAL-flag-MLH1*	2μ
pEAE296	*GAL-ha-PMS1*	2μ
pEAE298	*GAL-pms1Δ450-475*	2μ
pEAE299	*GAL-pms1Δ600-625*	2μ
pEAE300	*GAL-pms1Δ437-487*	2μ
pEAE301	*GAL-pms1Δ511-561*	2μ
pEAE302	*GAL-pms1Δ584-634*	2μ
pEAE303	*GAL-pms1Δ450-550*	2μ
pEAE304	*GAL-pms1Δ390-610*	2μ
pEAE308	*GAL-mlh1Δ348-373*	2μ
pEAE309	*GAL-mlh1Δ445-470*	2μ
pEAE310	*GAL-mlh1Δ359-409*	2μ
pEAE311	*GAL-mlh1Δ407-457*	2μ
pEAE312	*GAL-mlh1Δ357-457*	2μ
pEAE313	*GAL-mlh1Δ336-480*	2μ
pEAE314	*GAL-mlh1Δ396-421*	2μ
pEAI160	*mlh1Δ::KanMX4*	Integrating
pEAO36	*HIS3*	*ARS-CEN*
pEAO38	*LEU2*	*ARS-CEN*
pMH1	*GAL-MLH1-VMA1-CBD*	2μ
pMH8	*GAL-PMS1*	2μ

Table 3.2

TEV protease cleavage of *MLH1* confers a mutator phenotype *in vivo*

Genotype	*n*	Mutation rate (10^{-7}), (95% C.I.)		Relative to wild type	
		Sucrose	Galactose	Sucrose	Galactose
Wild type	15	4.9 (4.1-5.6)	7.8 (3.2-30)	1	1.6
mlh1Δ	15	15900 (10400-27700)	30100 (11900-55000)	3245	6143
flag-MLH1	15	4.7 (3.7-6.8)	76 (6.7-157)	1	16
tev-flag-MLH1	15	13 (9-122)	16700 (4860-37400)	2.7	3408
flag-tev-MLH1	15	5.6 (3.9-25)	28000 (13600-43300)	1.1	5714

The indicated alleles were integrated into the strain EAY2576 and were tested in the *lys2::insE-A*$_{14}$ mutator assay. TEV protease was under the galactose promoter allowing for carbon source-dependent cleavage of *MLH1* constructs containing consensus TEV protease cleavage sites within the unstructured linker arm region of the protein. Lys^+ reversion rates under uninduced (sucrose) and induced (galactose) conditions of TEV protease expression are indicated. *n* = number of independent measurements.

necessary due to the wild- type function of the TEV-tagged MLH1 constructs in the absence of TEV expression. Furthermore, this implies that TEV expression for the ~20 generations required to form a colony is too low to see maximal Lys^+ reversion whereas the ~60 generations that accrue before plating are sufficient. Overall, these results demonstrate that the DNA binding activity afforded by the MLH1-PMS1 linker arms is required for DNA mismatch repair.

A series of deletions in MLH1-PMS1 unstructured linker arm domains confers differential mismatch repair defects. Previous work characterizing MLH1-PMS1 DNA diffusion was consistent with a rapid hopping/stepping mechanism. Movement in this manner while potentially adopting an open ring-like configuration allowed efficient bypass of nucleosomes (Gorman *et al.* 2010). I reasoned that if the linker arm domains of MLH1-PMS1 were shortened then this would result in diminished diffusion along DNA and possibly an inability to bypass nucleosome barriers. To test this, a series of variable sized deletions were engineered and spaced out along the predicted linker arm domains of each protein (Figure 3.3). The limits of the linker arm domains were chosen conservatively to decrease the possibility of disrupting the N or C -terminal globular domains. *MLH1* alleles contain a FLAG-epitope tag in a position downstream of the conservative linker arm domain (after amino acid 499Y) that was previously shown to not disrupt MLH1 mismatch repair function (Argueso *et al.* 2003; Gorman *et al.* 2010). All *PMS1* alleles, except *pms1Δ390-610,* contain an HA-epitope tag within the linker arm domain (after amino acid 565D) that was also previously shown to be functional for mismatch repair. Epitope-tags were engineered to analyze expression levels (see below) and for future TIRFM experiments (see Discussion).

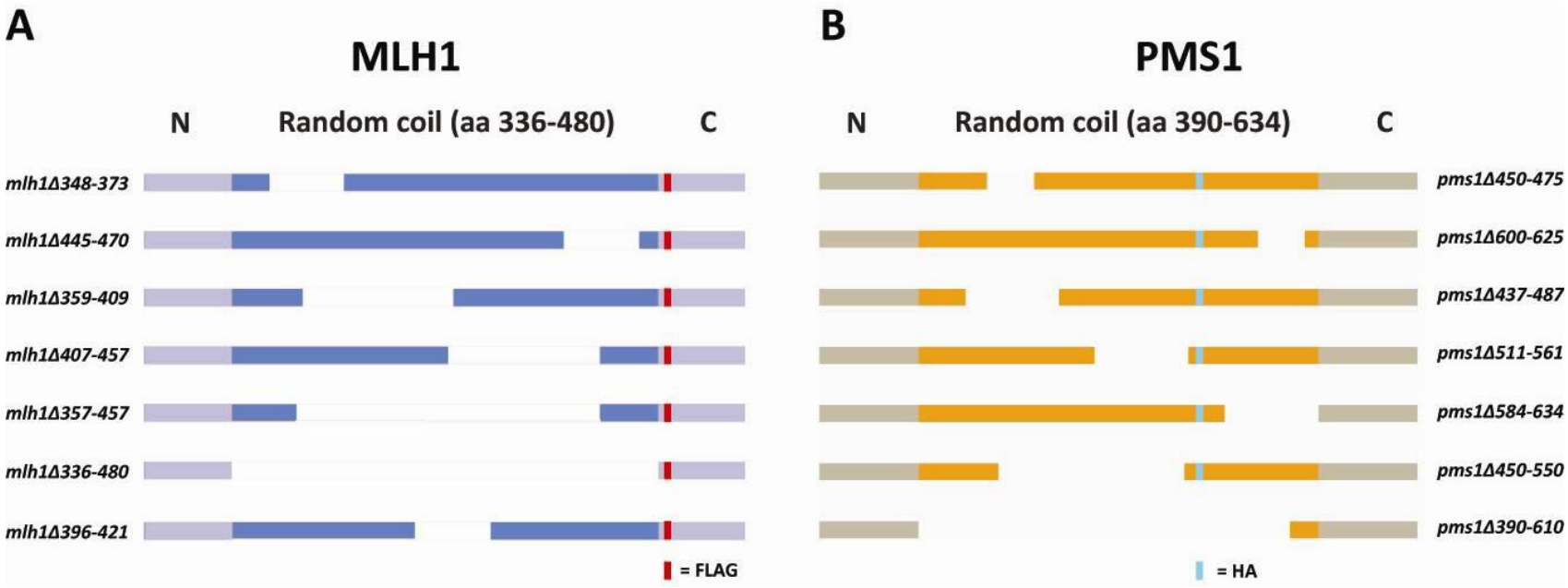


Figure 3.3. Schematic diagram of MLH1 and PMS1 linker arm deletion series.
(A) Series of amino acid deletions within the MLH1 linker arm domain with the relative size and position illustrated by gaps. Numbers in the mutant names indicate amino acids spanning the deletion. Location of the FLAG-epitope tag is illustrated by a red bar. **(B)** Similar series as in A, but for PMS1. Location of the HA-epitope tag is illustrated by a blue bar. Note that the N and C – termini of MLH1 and PMS1 are not drawn to scale.

Initially, individual linker domain mutants were tested in the *lys2*::*insE-A*$_{14}$ reversion assay in the presence of their wild-type heterodimeric partner (Table 3.3). Interestingly, the complete deletion of the linker arm in either *MLH1* or *PMS1* confers a null phenotype for MMR. With the exception of the 25 amino acid deletions *mlh1Δ348-373* and *mlh1Δ445-470,* all of the *mlh1* mutants displayed high mutation rates similar to an *mlh1Δ* strain. The exceptions were *mlh1Δ348-373* that displayed a moderately increased mutation rate (intermediate) and *mlh1Δ445-470* that was indistinguishable from wild-type. In contrast, most of the *pms1* linker domain mutants were indistinguishable from wild-type. A 25 amino acid deletion *pms1Δ600-625* and a 50 amino acid deletion *pms1Δ584-634* displayed mutation rates at an intermediate level and the complete deletion *pms1Δ390-610* was indistinguishable from a *pms1Δ* strain. These results indicate that the MLH1 linker arm is more sensitive to shortening compared to that of PMS1.

In order to look for synthetic defects, mutant alleles of *mlh1* and *pms1* that were either wild-type or intermediate in the *lys2*::*insE-A*$_{14}$ reversion assay were retested in partnered combinations (Table 3.3). Combinations with *mlh1Δ445-470* recapitulated the mutator phenotype of the individual *pms1* allele. This was not surprising since *mlh1Δ445-470* alone was indistinguishable from wild-type in this assay. Interestingly, combinations with the intermediate allele *mlh1Δ348-373* increased the mutation rate to a level that was multiplicative from the individual rates. This reveals that the mild defect in the individual mutants is severely exacerbated when combined with a partner that has a shortened linker arm. I interpret this to mean that shortening of the linker arm domains of MLH1-PMS1 causes a defect in mismatch repair that is possibly due to a reduction in the diffusion rate along DNA or an inability to bypass nucleosome barriers. Future TIRFM studies are necessary to test this

Table 3.3

mlh1 and _pms1_ linker arm deletions confer differential mutator phenotypes

Genotype	*n*	Mutation Rate (10^{-7}), (95% C.I.)	Relative to wild type
Wild type (*MLH1*)	20	7.5 (3.5–18)	1
FLAG-MLH1	20	7.7 (5.2-25)	1
mlh1Δ	20	45,100 (23,000-255,000)	5,997
mlh1Δ396-421	20	89,700 (15,900-180,000)	11,960
mlh1Δ348-373	20	323 (132-1,080)	43
mlh1Δ445-470	20	7.5 (3.3-13)	1
mlh1Δ359-409	20	15,800 (10,000-37,400)	2,107
mlh1Δ407-457	20	49,400 (14,700-127,000)	6,587
mlh1Δ357-457	20	49,500 (11,300-169,000)	6,600
mlh1Δ336-480	20	53,000 (22,900-70,600)	7,067
Wild type (*PMS1*)	15	1.5 (0.9-2.4)	1
HA-PMS1	15	5.3 (4.1-17)	3.5
pms1Δ	15	23,100 (14,000-76,100)	15,400
pms1Δ450-475	15	7.1 (5.2-9.0)	4.7
pms1Δ600-625	15	489 (127-916)	326
pms1Δ437-487	15	9.5 (6.6-21)	6.3
pms1Δ511-561	15	16 (6.9-24)	10.7
pms1Δ584-634	15	415 (152-839)	277
pms1Δ450-550	15	10 (6.4-84)	6.7
pms1Δ390-610	15	14,300 (8,230-23,000)	9,533

Wild type (*MLH1*, *PMS1*)	15	2.1 (0.8-5.8)	1
FLAG-MLH1, *HA-PMS1*	15	14 (8-28)	6.7
mlh1Δ pms1Δ	15	13,800 (10,800-26,000)	6,571
mlh1Δ348-373, *pms1Δ450-475*	15	563 (195-928)	268
mlh1Δ348-373, *pms1Δ600-625*	15	16,100 (3,440-137,000)	7,667
mlh1Δ348-373, *pms1Δ437-487*	15	511 (407-658)	243
mlh1Δ348-373, *pms1Δ511-561*	15	917 (494-1,770)	437
mlh1Δ348-373, *pms1Δ584-634*	15	6,760 (4,780-11,800)	3,219
mlh1Δ348-373, *pms1Δ450-550*	15	835 (585-1,740)	398
mlh1Δ445-470, *pms1Δ450-475*	15	7.9 (4.5-31)	3.8
mlh1Δ445-470, *pms1Δ600-625*	15	285 (135-672)	136
mlh1Δ445-470, *pms1Δ437-487*	15	8.0 (6.7-16)	3.8
mlh1Δ445-470, *pms1Δ511-561*	15	22 (12-47)	10.5
mlh1Δ445-470, *pms1Δ584-634*	15	205 (132-343)	98
mlh1Δ445-470, *pms1Δ450-550*	15	25 (10-77)	12

mlh1 and *pms1* alleles listed were tested in the *lys2::insE-A*$_{14}$ mutator assay. For each allele the amino acids spanning the deletion is indicated. *Lys+* reversion rates were calculated for *mlh1*, *pms1*, and combinations of alleles in strains EAY1366, EAY3097, and EAY1365, respectively (see Table 3.1). For each strain, the allele was expressed from an *ARS-CEN* plasmid under the native promoter of the corresponding wild-type gene. *n* = number of independent measurements

hypothesis.

MLH1-PMS1 linker arm deletion mutants can be expressed and purified. Individual *mlh1* and *pms1* mutant alleles were over-expressed under the galactose inducible promoter to assess protein stability. After induction, crude extracts were collected and mutant proteins were identified by Western blot analysis using antibodies against the relevant epitope-tags. Using an anti-FLAG antibody, each mlh1 linker arm deletion polypeptide was detected at levels equivalent to those seen in extracts containing an otherwise wild-type FLAG-tagged MLH1 protein (Figure 3.4A). pms1 linker arm deletions polypeptides were detected using an anti-HA antibody (Figure 3.4B). *pms1Δ390-610* expression levels cannot be tested in this manner because it does not contain an HA-tag. *pms1Δ600-625* had severely reduced expression levels compared to HA-PMS1. This result may explain the intermediate mutator phenotype seen for this allele in the *lys2*::*insE-A*$_{14}$ reversion assay. Conversely, the other *pms1* mutants were detected at levels similar to those seen in a strain containing wild-type HA-tagged PMS1 protein. Taken together the protein expression results indicate that the defect in mismatch repair caused by MLH1-PMS1 linker arm deletions is not due to a lack of protein expression or stability.

In order to characterize the defect caused by shortened linker arms, interesting candidate complexes were purified from the same strains used in the expression experiments above. Purification was performed as previously described (Hall and Kunkel 2001); Chapter 2). Initially, I attempted to purify the largest combined deletion complex (*mlh1Δ336-480*, *pms1Δ390-610*) because this complex would be the most likely to show an observable defect in bulk and TIRFM assays. Unfortunately, for reasons that are unknown purification of this

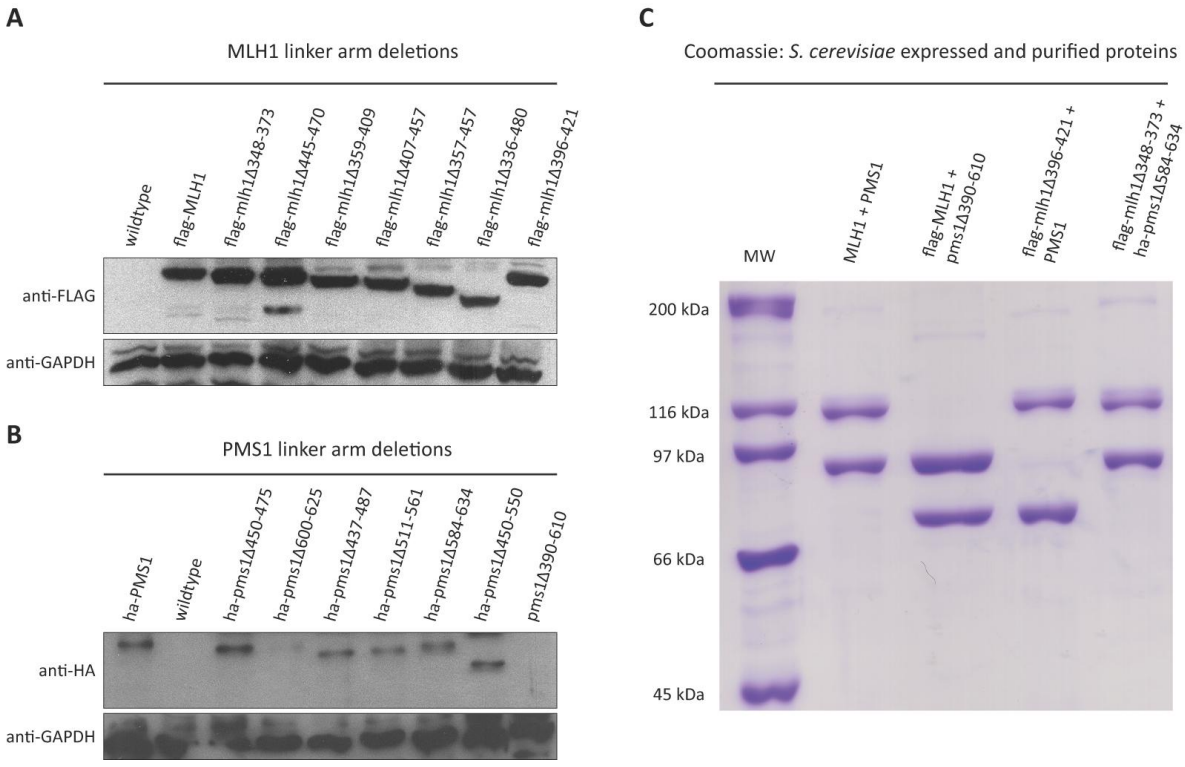


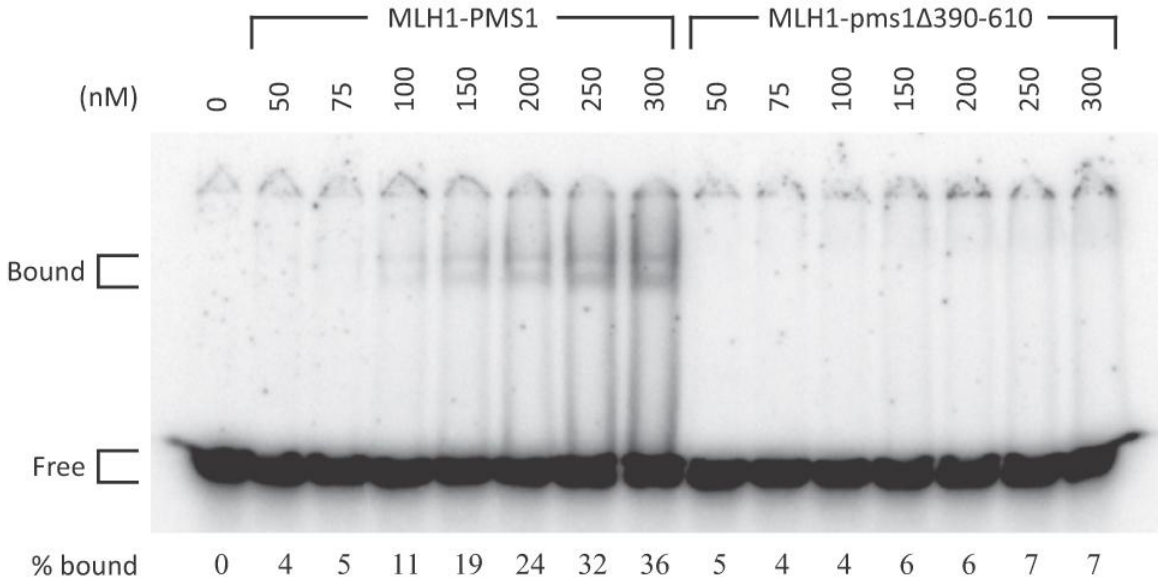
Figure 3.4. MLH1-PMS1 linker arm deletion mutants can be expressed and purified. **(A)** Crude cellular extracts from strains bearing the indicated *MLH1* allele were separated by 8% SDS-PAGE and then probed with an anti-FLAG antibody. **(B)** Crude cellular extracts from strains bearing the indicated *PMS1* allele were separated by 8% SDS-PAGE and then probed with an anti-HA antibody. Anti-GAPDH was used as a loading control for both **(A)** and **(B)**. **(C)** Indicated combination of MLH1 and PMS1 wild-type or linker arm deletion proteins were expressed and purified from *S. cerevisiae* (see methods). 0.5 μg of protein loaded into each lane was stained with Coomassie blue after separation by 8% SDS-PAGE. The sizes of the relevant molecular weight standards are indicated.

complex was unsuccessful (data not shown). In light of this, individual complete linker arm deletions of MLH1 (*mlh1Δ336-480*) and PMS1 (*pms1Δ390-610*) were purified with their wild-type partner (Figure 3.4C). Purification of MLH1-pms1Δ390-610 confirms that untagged *pms1Δ390-610* is stably expressed. The synthetically null complex *mlh1Δ348-373, pms1Δ584-634* was also chosen for purification to analyze the nature of the combination-dependent defect. Due to the fact that these mutant complexes can be purified, I conclude that formation of an MLH1-PMS1 heterodimeric complex does not require intact linker arm domains.

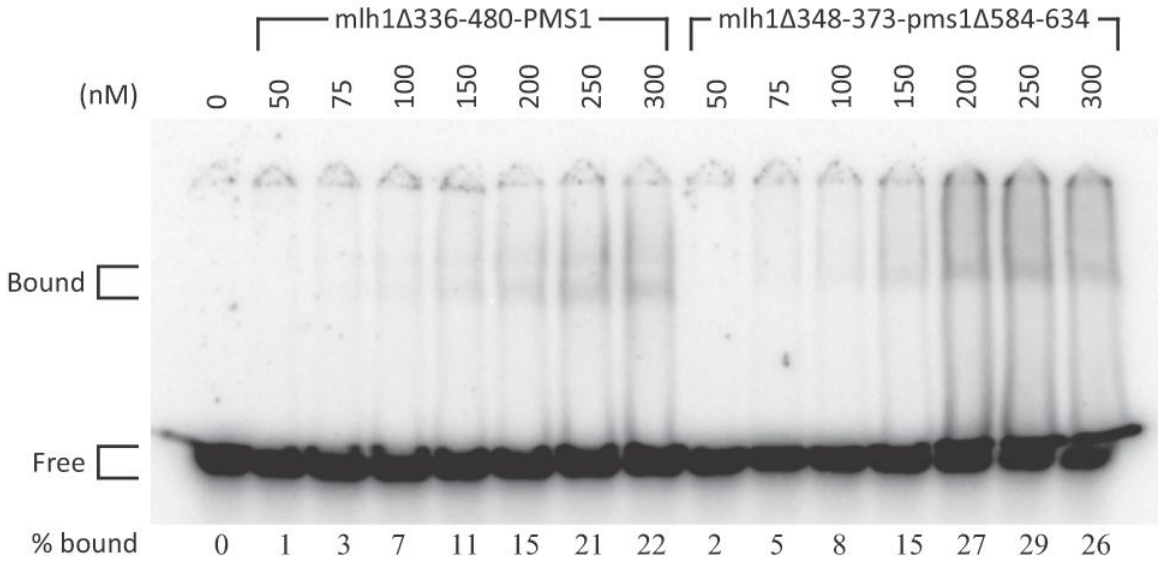
Deletions of MLH1-PMS1 linker arms have disparate effects on DNA binding activity and association with MSH2-MSH6 at a mismatch. MLH1-PMS1 binds to DNA through non-specific backbone contacts with no specificity for DNA mismatches (Hall *et al.* 2001; Gorman *et al.* 2010). The mismatch repair defect that was observed in the genetic assay for the different linker arm deletion complexes may be due to an impairment of stable DNA binding activity or a disruption of the association of MLH1-PMS1 with MSH2-MSH6 bound to a mismatch. In order to test if purified mlh1-pms1 complexes are able to interact with DNA, I carried out electromobility shift assays (EMSA) with short (40-bp) radio-labeled oligonucleotides (Figure 3.5A, B). Two of the three mutant complexes analyzed (mlh1Δ336-480-PMS1 and mlh1Δ348-373-pms1Δ584-634) shifted the DNA substrate within the protein concentration range tested. DNA binding by MLH1-pms1Δ390-610 was not detected at protein concentrations up to 500 nM (data not shown). Apparent dissociation constants were determined in titration experiments as the protein concentration at 50% maximal binding (Figure 3.5C). The complete deletion of the MLH1 linker arm complexed with PMS1

Figure 3.5. *mlh1* and *pms1* linker arm deletions display a range of altered DNA binding affinities.
EMSA was performed as described in Materials and Methods. All reactions contained 60nM 40-bp homoduplex substrate. **(A)** Titration reactions containing the indicated amounts of MLH1-PMS1 or MLH1-pms1Δ390-610 complexes. Free and bound substrates are indicated by brackets. % bound was calculated using ImageJ software as the amount bound divided by the total (bound + free) and is indicated below each lane. **(B)** Titration reaction as for **(A)** but with mlh1Δ336-480-PMS1 or mlh1Δ348-373-pms1Δ584-634 complexes. **(C)** Apparent Kd values calculated from experiments in **(A)** and **(B)**. The values represent the concentration at half maximal binding.

A



B



C

Apparent Kd values	
MLH1-PMS1	395 nM
MLH1-pms1Δ390-610	4100 nM
mlh1Δ336-480-PMS1	592 nM
mlh1Δ348-373-pms1Δ584-634	461 nM

(mlh1Δ336-480-PMS1) displayed a DNA binding affinity that was only slightly less than the wild-type MLH1-PMS1 complex (~67% of wt). Likewise, the synthetically null complex (mlh1Δ348-373-pms1Δ584-634) had a binding affinity near wild-type levels (~86% of wt). From these results, I conclude that the defect caused by these complexes is not due to a lack of an association with DNA. In contrast, the complete deletion of the PMS1 linker arm complexed with MLH1 (MLH1-pms1Δ390-610) had an affinity for DNA that was more than an order of magnitude greater than wild-type. This suggests that the PMS1 linker arm may be more critical for DNA binding and less amenable to shortening.

Interactions between MLH1-PMS1 and MSH2-MSH6 bound to a mismatch can be observed by EMSA where the two complexes generate a super-shifted species. This interaction requires ATP and is thought to form in order to signal to the downstream effectors to complete repair (Habraken *et al.* 1998). Although the conserved connector domain II of MSH2-MSH6 has been shown to be important for interaction between MutS and MutL, no information is currently present about the region of MLH1-PMS1 that is required for this association (Mendillo *et al.* 2009). Ternary complex formation through EMSA was used to rule out the possibility that deletions within the linker domain of MLH1 and PMS1 disrupted association with MSH2-MSH6 at a mismatch. MLH1-PMS1 complexes were present at a concentration below detectable independent DNA binding to remove the possibility of blocking MSH2-MSH6 access to the mismatch site. Mutant complexes that I previously observed to have detectable DNA binding activity (mlh1Δ336-480-PMS1 and mlh1Δ348-373-pms1Δ584-634) were also able to interact with MSH2-MSH6 at a mismatch (Figure 3.6). Interestingly, MLH1-PMS1 DNA binding activity was stimulated by MSH2-MSH6 to create the two distinct band-shifts characteristic of MLH1-PMS1, as well as the super-shifted ternary

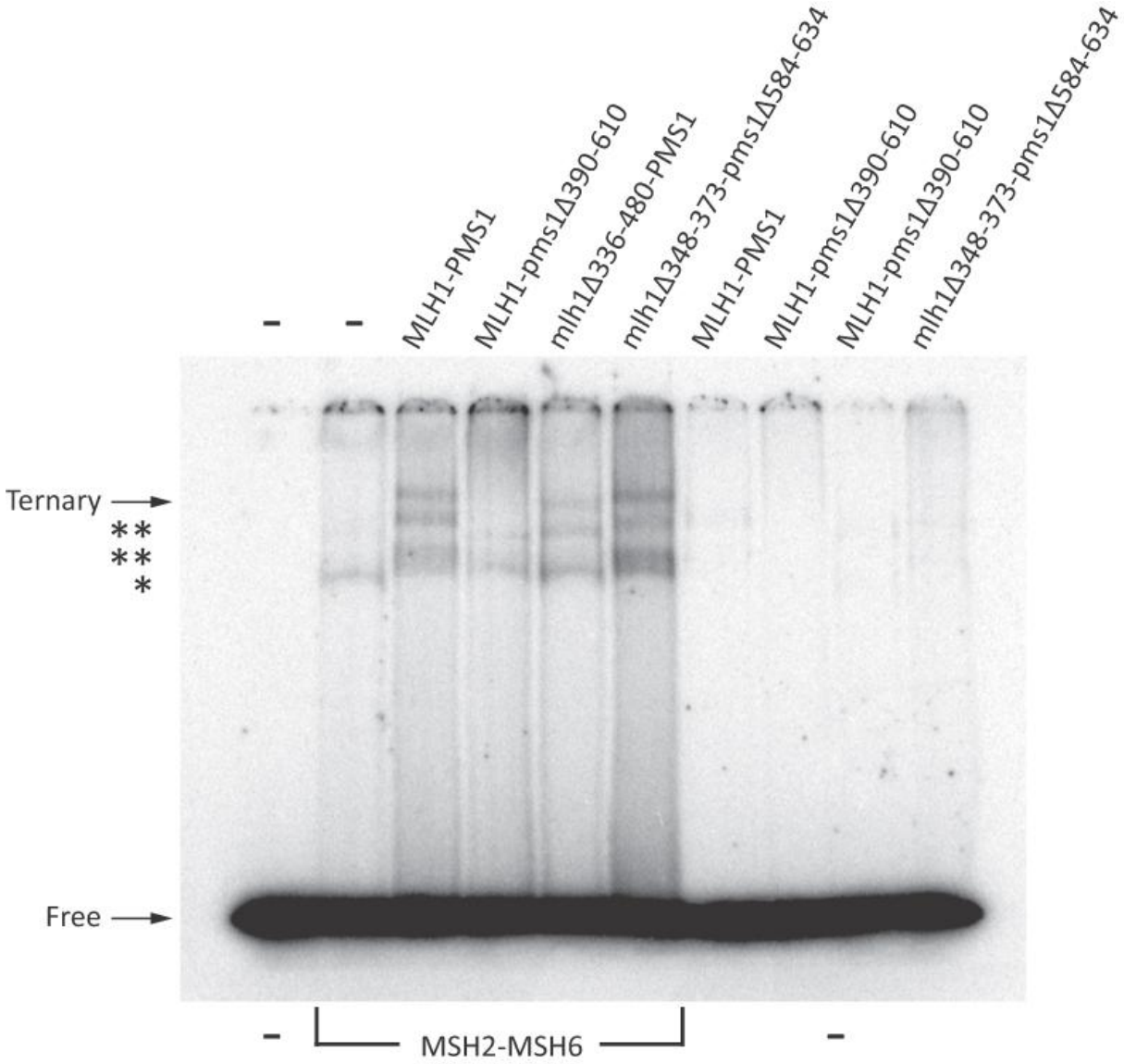


Figure 3.6. *mlh1* and *pms1* linker arm deletion mutants that are able to bind to DNA are also able to form a ternary complex with MSH2-MSH6 and a mismatch.
EMSA was performed as described in Materials and Methods. All reactions contained 60nM 40bp (+1) mismatch substrate. 150nM MSH2-MSH6 was present in lanes that are bracketed. Indicated MLH1-PMS1 complexes are present at 100nM. (-) at the bottom of gel indicate no MSH2-MSH6 present and (-) at the top of gel indicates no MLH1-PMS1 present. Free substrate and shifted ternary complexes are indicated by arrows. * - indicates MSH2-MSH6 gel-shift, ** - indicates MLH1-PMS1 gel-shift. Note: MLH1-PMS1 gel-shift results in two different sized shifted species (see Figure 3.5).

complex. This result may be explained by the fact that MLH1-PMS1 has been shown to display cooperative binding and the presence of MSH2-MSH6 on the short oligonucleotide may facilitate DNA association (Hall *et al.* 2001). MLH1-pms1Δ390-610 was not able to form a ternary complex suggesting that DNA-binding is a prerequisite for association with MSH2-MSH6.

Discussion and Future Directions

In this study I show that the unstructured linker arms of MLH1 and PMS1 are important for the complex to bind DNA and carry out mismatch repair functions. Proteolytic cleavage with TEV protease and EMSA analysis showed that an intact linker arm was required for stable DNA binding. Using the *lys2*::*insE-A*$_{14}$ assay, I showed that TEV cleavage of the linker arm in MLH1 resulted in a high mutator phenotype *in vivo*. Using the same mismatch repair assay, I characterized the phenotypes of a series of deletions within the linker arms of *MLH1* and *PMS1*. This analysis revealed that the linker arm of MLH1 is more sensitive to deletion, in that, even a relatively small 25 amino acid deletion caused a high mutator phenotype. In contrast, removal of up to 100 amino acids in the linker arm of PMS1 had no effect on mismatch repair function. Combinatorial analysis with alleles that had intermediate or no phenotype in the *lys2*::*insE-A*$_{14}$ assay revealed complexes with a defect that was the product of the individual alleles. Purified linker arm deletion complexes displayed a range of DNA binding affinities from near wild-type for mlh1Δ348-373- pms1Δ584-634, slightly reduced affinity for mlh1Δ336-480-PMS1, to a severe defect in DNA binding for MLH1- pms1Δ390-610. I found that MLH1-PMS1 complexes that could bind DNA were also competent to form a ternary complex with MSH2-MSH6 at a mismatch.

MLH1-PMS1 forms a ring-like structure that encircles DNA. MLH1-PMS1 has been hypothesized to form a ring-like structure capable of encircling DNA. In support of this, experiments have been done that reveal in the absence of DNA ends MLH1-PMS1 becomes trapped on the DNA but when it encounters a free DNA end it dissociates at a much higher frequency than at any internal position (Gorman *et al.* 2010); Chapter 2). In this work, I observed loss of DNA binding by MLH1-PMS1 and loss of MMR when a break is made in the linker arms of MLH1-PMS1 by TEV protease. Contrary to what I found, studies with isolated NTDs from MLH1 and PMS1, as well as, isolated CTDs from *E. coli* MutL have been shown to bind stably to DNA (Arana *et al.*; Guarne *et al.* 2001; Hall *et al.* 2003). I believe the differences I saw in the TEV experiments are due to residual associations between the cleaved fragments that are inhibiting DNA binding through an unknown mechanism. In support of this, co-immunoprecipitation experiments done after TEV cleavage suggests that the two resulting fragments of MLH1 can still interact with full-length PMS1 (data not shown). Functional communication between the two domains is known to be important for ATP-dependent N-terminal associations and in the context of full-length MLH1-PMS1 may be required for coordination of DNA binding (Guarne *et al.* 2001; Hall *et al.* 2003). The fact that MLH1-PMS1 has multiple DNA binding sites that map to both subunits might necessitate such coordination (Hall *et al.* 2003). Nonetheless, the finding that TEV cleavage of the linker arm of MLH1 in yeast cells resulted in an elevated mutation rate implies that the DNA binding activity of the intact heterodimer is important for mismatch correction.

A role for the unstructured linker arms of MLH1-PMS1 in diffusion along DNA?

Deletion analysis along the linker arms of MLH1-PMS1 revealed that the length and integrity

of MLH1 was more critical for mismatch repair activity. My findings for PMS1 are similar to those seen when truncations were made in the linker arm of *E. coli* MutL (Guarne *et al.* 2004). In that study they found that deletions up to one-third the size of the linker arm in MutL did not disrupt DNA binding activity or MMR function. The fact that the linker arm in PMS1 is twice the size of MLH1 may allow for larger truncations to be made without effect. Alternatively, there may be important residues along the MLH1 linker arm that are critical for protein function. In support of this, an alanine-scanning mutagenesis screen identified a mismatch repair defective allele, *mlh1-31,* that is mutated at residues R401 and D403 and overlaps with several of my non-functional deletion constructs (Argueso *et al.* 2003). Interestingly, mlh1-31 was still able to associate with PMS1 and form a ternary complex with MSH2-MSH6 at a mismatch suggesting it has a defect in downstream repair functions that are possibly associated with other protein-protein interactions. It is also worth mentioning that not all of my deletion constructs that had a null-phenotype removed these key residues.

Expression analysis ruled out the possibility that any null-phenotype seen for the linker arm deletions constructs was a consequence of protein instability or lack of protein expression. This finding allowed me to purify and test interesting complexes for DNA binding activity and MSH2-MSH6 interactions. To narrow my options I reasoned that the largest disruption in the complex (mlh1Δ336-480-pms1Δ390-610) would likely show interesting behaviors in bulk and TIRFM assays. For reasons that are unknown this complex was refractory to purification in significant quantities. A confounding problem is that, after complete deletion of the linker arms, both subunits were roughly the same size and difficult to discern by SDS-PAGE. Purification of the complete deletion in individual subunit linker arms in combination with the full-length partner was successful. These complexes displayed

differences in their ability to bind DNA and interact with MSH2-MSH6. In contrast to my results for mismatch repair function, it appears that the linker arm of PMS1 is more important for the DNA binding function of this complex. Likewise, deletion of the linker arm in PMS1 impaired the ability to interact with MSH2-MSH6 but the linker arm deletion in MLH1 could still form a ternary complex.

Combinatorial analysis of wild-type and intermediate linker arm deletions revealed a multiplicative effect between some complexes. This effect was only seen with the intermediate *mlh1* allele (*mlh1Δ348-373*), whereas combinations with the only wild-type *mlh1* allele (*mlh1Δ445-47*) displayed the individual phenotype of the *pms1* allele tested. These results indicate that combinations of weakened alleles have drastic effects on protein function. One possibility is that there is a threshold size required for the complex to form a functional ring and that this is being surpassed in my defective combinations. Another possibility is that the hopping/stepping diffusion mechanism used by this complex may rely on the linker arms acting in a manner similar to arms or legs and when they are shortened this may impair diffusion and obstacle bypass. In the future it will be important to address if these complexes have defects in diffusion along DNA or an inability to bypass nucleosomes by TIRFM. This will allow conclusions to be drawn about the importance of such activity on the mismatch repair function for this complex. Importantly, the synthetic null complex that I purified was able to bind DNA and form a ternary complex similar to wild-type MLH1-PMS1.

Mechanistic details of the MLH1-PMS1 complex as it diffuses on DNA will aid in providing contextual information to the known functions of this complex. It has been suspected that conformational changes involving the linker arms of MLH1-PMS1 may promote essential interactions with other MMR components (Sacho *et al.* 2008). Interactions

with PCNA and Exo1 are of particular importance as they may be necessary to stimulate and complete the excision step of MMR (Tran *et al.* 2001; Lee and Alani 2006). MLH1-PMS1 has been shown to harbor a latent endonuclease activity that is attributed to the C-terminal domain of PMS1 (Kadyrov *et al.* 2006). Activation of this activity is also thought to be triggered by interactions with other mismatch repair components in order to displace a regulatory subdomain that blocks access to DNA (Pillon *et al.* 2010). In the future it will be interesting to see how the linker arms of MLH1-PMS1 are coordinating diffusion along DNA with the sequential protein-protein interactions steps of MMR.

Materials and Methods

Strains and Plasmids. The plasmids and strains used in this study are listed in Table 3.1. Full details of plasmid and strain constructions available upon request.

Linker arm deletion series construction. For each linker arm deletion separate vectors were created for over-expression and for complementation tests. *mlh1* linker arm deletions complementation vectors were constructed as derivatives of the *ARS-CEN LEU2* vector, pEAA213, expressing *MLH1* from the native *MLH1* promoter (Heck *et al.* 2006). *pms1* linker arm deletions complementation vectors were constructed as derivatives of the *ARS-CEN HIS3* vector, pEAA238, expressing *PMS1* from the native *PMS1* promoter (Heck *et al.* 2006). Expression vectors were constructed as derivates of pMH1 (*GAL1-MLH1-VMA-CBD*, 2μ, *TRP1*) and pMH8 (*GAL10-PMS1*, 2μ, *LEU2*) for *mlh1* and *pms1* linker arm deletions, respectively (Hall and Kunkel 2001). Each deletion was constructed by overlap-extension

PCR to remove the portion of the corresponding protein, as indicated (Ho *et al.* 1989). DNA fragments containing the relevant linker arm deletions inserted into pEAA213, pEAA238, pMH1, and pMH8 were confirmed by DNA sequencing (Cornell BioResource Center).

Electromobility shift assays. Protein concentrations were determined by Bradford assay using BSA as a standard. Gel mobility shift assays with oligonucleotide substrates were performed as described (Kijas *et al.* 2003). Briefly, MLH1-PMS1 titration binding reactions were assembled on ice in 15 µl reactions containing 60 nM (5'-^{32}P)-end labeled 40-bp homoduplex substrate, 25 mM Hepes [pH, 7.6], 40 µg/ml BSA, 1 mM DTT, 50 mM NaCl, and 8% Sucrose (w/v). MLH1-PMS1 constructs (0-300nM) were added last followed by a 5 minute incubation at room temperature (RT). In ternary complex assays, reactions contained 1 mM ATP, 150 nM MSH2-MSH6, 100 nM of indicated MLH1-PMS1 constructs and were incubated with 60 nM (5'-^{32}P)-end labeled 40-bp (+1) mismatch substrate. After incubation, samples were loaded on 4% (w/v) non-denaturing polyacrylamide gels containing 0.5X TBE and electrophoresed at 130 V for 1 hour at RT. Gels were dried on 3MM Whatman paper and visualized by PhosphorImaging. Kinetic analysis was done using ImageJ. The 40-bp substrate homoduplex substrate was created by annealing S1 (5'dACCGAATTCTGACTTGCTAGGACATCTTTGCCCACGTTGA) and S2 (5'dTCAACGTGGGCAAAGATGTCCTAGCAAGTCAGAATTCGGT) and the 40-bp (+1) mismatch substrate was created by annealing S6 (5'dACCGAATTCTGACTTGCTAGAGACATCTTTGCCCACGTTGA) and S2 (Integrated DNA Technologies; (Surtees and Alani 2006).

TEV cleavage assays. 3 μg of MLH1-PMS1 was incubated with 0.03 μg of TEV protease in 15 μl reactions containing 25 mM Hepes [pH 7.6], 1 mM DTT, and 40 μg/ml BSA. TEV protease was a generous gift from Dr. Ailong Ke (Cornell). TEV cleaved proteins were assayed for DNA binding activity using gel shift assays, as described above (Figure 3.2). To confirm TEV cleavage, samples were incubated at 30°C for 30 minutes, after which 7.5 μl of 3X SDS-loading buffer (0.195 M Tris [pH 6.8], 30% glycerol, 3% β-mercaptoethanol, 6% SDS) was added to each and samples were boiled for 3 minutes. Samples were analyzed by 10% SDS-PAGE and stained with Coomassie blue (Figure 3.2) (Gorman *et al*. 2010; Chapter 2) .

***lys2*::*insE-A*$_{14}$ reversion assay.** pEAA213 (*MLH1*) and pEAA238 (*PMS1*) and derivative plasmids were transformed into EAY1366 (*mlh1Δ, lys2*::*insE-A*$_{14}$ *)* and EAY3097 (*pms1Δ, lys2*::*insE-A*$_{14}$), respectively, using standard methods and were maintained on minimal histidine, leucine dropout plates. When tested in combination, plasmids were transformed into EAY1365 (*mlh1Δ pms1Δ, lys2*::*insE-A*$_{14}$) as described above. *In vivo* TEV assays were preformed in strains EAY3098-EAY3102 that were maintained on minimal media containing either 4% sucrose as the sole carbon source or 2% Sucrose + 2% Galactose as carbon sources. Each strain was sequenced to confirm integrations and to verify *lys2*::*insE-A*$_{14}$ integrity. Rates of *lys2*::*insE-A*$_{14}$ reversion were calculated as $\mu = f/\ln(N\cdot\mu)$, where f is reversion frequency and N is the total number of revertants in the culture (Tran *et al.* 1997). 95% confidence intervals and all computer aided rate calculations were performed as previously described (Demogines *et al.* 2008).

Protein expression and purification. MLH1–PMS1 was expressed and purified using pMH1 (*GAL1-MLH1-VMA-CBD*, 2μ, *TRP1*) and pMH8 (*GAL10-PMS1*, 2μ, *LEU2*) vectors transformed into the *S. cerevisiae* strain BJ2168 (Hall *et al.* 2001); Chapter 2). MLH1–PMS1 linker arm deletion complexes were purified from BJ2168 containing the relevant pMH1 and pMH8 derivatives described above. Western blot analysis was performed on cell lysates collected after galactose induction. Cells were pelleted, washed with chitin buffer (25mM Tris, pH 8.0, 500mM NaCl, 10% glycerol, 1mM EDTA), repelleted, and resuspended in SDS–protein loading buffer. Samples were subjected to 8% SDS–PAGE and then transferred to a nitrocellulose membrane. Membranes were blocked with 4% milk overnight and probed with a 1:2000 dilution of 12CA5 (αHA, Roche) or a 1:1000 dilution of M2 (αFLAG, Sigma) antibody, followed by incubation with a 1:5000 dilution of α-mouse IgG secondary antibody (Jackson Immunoresearch). Proteins were visualized by the ECL detection method (Amersham/GE).

Acknowledgements

I would like to thank members of the Alani laboratory for all their help and support. I am grateful to Alba Guarne for bioinformatics help on the limits of the unstructured linker arm domains of MLH1-PMS1. I am very appreciative of the TEV-integrated strain EAY 2576 that was provided by Kim Nasmyth. Also, thank you to Ailong Ke for the kind gift of TEV protease, the Cornell R3 group for helpful comments and fruitful discussions, and Pei Xin Lim for construction of EAY 3097. This project is being conducted in collaboration with Eric Greene's lab at Columbia University.

References

Arana, M. E., S. F. Holmes, et al. "Functional residues on the surface of the N-terminal domain of yeast Pms1." DNA Repair (Amst) **9**(4): 448-57.

Argueso, J. L., A. W. Kijas, et al. (2003). "Systematic mutagenesis of the Saccharomyces cerevisiae MLH1 gene reveals distinct roles for Mlh1p in meiotic crossing over and in vegetative and meiotic mismatch repair." Mol Cell Biol **23**(3): 873-86.

Ban, C., M. Junop, et al. (1999). "Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair." Cell **97**(1): 85-97.

Ban, C. and W. Yang (1998). "Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis." Cell **95**(4): 541-52.

Berg, O. G., R. B. Winter, et al. (1981). "Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory." Biochemistry **20**(24): 6929-48.

Clark, A. B., F. Valle, et al. (2000). "Functional interaction of proliferating cell nuclear antigen with MSH2-MSH6 and MSH2-MSH3 complexes." J Biol Chem **275**(47): 36498-501.

Demogines, A., A. Wong, et al. (2008). "Incompatibilities involving yeast mismatch repair genes: a role for genetic modifiers and implications for disease penetrance and variation in genomic mutation rates." PLoS Genet **4**(6): e1000103.

Dutta, R. and M. Inouye (2000). "GHKL, an emergent ATPase/kinase superfamily." Trends Biochem Sci **25**(1): 24-8.

Flores-Rozas, H., D. Clark, et al. (2000). "Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mispair recognition complex." Nat Genet **26**(3): 375-8.

Gorman, J., A. Chowdhury, et al. (2007). "Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6." Mol Cell **28**(3): 359-70.

Gorman, J., A. J. Plys, et al. (2010). "Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice." Nat Struct Mol Biol **17**(8): 932-8.

Gradia, S., S. Acharya, et al. (1997). "The human mismatch recognition complex hMSH2-hMSH6 functions as a novel molecular switch." Cell **91**(7): 995-1005.

Gradia, S., D. Subramanian, et al. (1999). "hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA." Mol Cell **3**(2): 255-61.

Groth, A., W. Rocha, et al. (2007). "Chromatin challenges during DNA replication and repair." Cell **128**(4): 721-33.

Guarne, A., M. S. Junop, et al. (2001). "Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase." Embo J **20**(19): 5521-31.

Guarne, A., S. Ramon-Maiques, et al. (2004). "Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair." Embo J **23**(21): 4134-45.

Habraken, Y., P. Sung, et al. (1998). "ATP-dependent assembly of a ternary complex consisting of a DNA mismatch and the yeast MSH2-MSH6 and MLH1-PMS1 protein complexes." J Biol Chem **273**(16): 9837-41.

Halford, S. E. and J. F. Marko (2004). "How do site-specific DNA-binding proteins find their targets?" Nucleic Acids Res **32**(10): 3040-52.

Hall, M. C. and T. A. Kunkel (2001). "Purification of eukaryotic MutL homologs from Saccharomyces cerevisiae using self-affinity technology." Protein Expr Purif **21**(2): 333-42.

Hall, M. C., P. V. Shcherbakova, et al. (2003). "DNA binding by yeast Mlh1 and Pms1: implications for DNA mismatch repair." Nucleic Acids Res **31**(8): 2025-34.

Hall, M. C., H. Wang, et al. (2001). "High affinity cooperative DNA binding by the yeast Mlh1-Pms1 heterodimer." J Mol Biol **312**(4): 637-47.

Heck, J. A., J. L. Argueso, et al. (2006). "Negative epistasis between natural variants of the Saccharomyces cerevisiae MLH1 and PMS1 genes results in a defect in mismatch repair." Proc Natl Acad Sci U S A **103**(9): 3256-61.

Ho, S. N., H. D. Hunt, et al. (1989). "Site-directed mutagenesis by overlap extension using the polymerase chain reaction." Gene **77**(1): 51-9.

Kadyrov, F. A., L. Dzantiev, et al. (2006). "Endonucleolytic function of MutLalpha in human mismatch repair." Cell **126**(2): 297-308.

Kijas, A. W., B. Studamire, et al. (2003). "Msh2 separation of function mutations confer defects in the initiation steps of mismatch repair." J Mol Biol **331**(1): 123-38.

Kohler, F. (2003). "A yeast-based growth assay for the analysis of site-specific proteases." Nucleic Acids Res **31**(4): e16.

Kosinski, J., I. Steindorf, et al. (2005). "Analysis of the quaternary structure of the MutL C-terminal domain." J Mol Biol **351**(4): 895-909.

Kunkel, T. A. and D. A. Erie (2005). "DNA mismatch repair." Annu Rev Biochem **74**: 681-710.

Ladoux, B., J. P. Quivy, et al. (2000). "Fast kinetics of chromatin assembly revealed by single-molecule videomicroscopy and scanning force microscopy." Proc Natl Acad Sci U S A **97**(26): 14251-6.

Lau, P. J. and R. D. Kolodner (2003). "Transfer of the MSH2.MSH6 complex from proliferating cell nuclear antigen to mispaired bases in DNA." J Biol Chem **278**(1): 14-7.

Lee, S. D. and E. Alani (2006). "Analysis of interactions between mismatch repair initiation factors and the replication processivity factor PCNA." J Mol Biol **355**(2): 175-84.

Li, F., L. Tian, et al. (2009). "Evidence that nucleosomes inhibit mismatch repair in eukaryotic cells." J Biol Chem **284**(48): 33056-61.

Lynch, H. T., P. M. Lynch, et al. (2009). "Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications." Clin Genet **76**(1): 1-18.

Mendillo, M. L., V. V. Hargreaves, et al. (2009). "A conserved MutS homolog connector domain interface interacts with MutL homologs." Proc Natl Acad Sci U S A **106**(52): 22223-8.

Modrich, P. (2006). "Mechanisms in eukaryotic mismatch repair." J Biol Chem **281**(41): 30305-9.

Modrich, P. and R. Lahue (1996). "Mismatch repair in replication fidelity, genetic recombination, and cancer biology." Annu Rev Biochem **65**: 101-33.

Pillon, M. C., J. J. Lorenowicz, et al. (2010). "Structure of the endonuclease domain of MutL: unlicensed to cut." Mol Cell **39**(1): 145-51.

Riggs, A. D., S. Bourgeois, et al. (1970). "The lac repressor-operator interaction. 3. Kinetic studies." J Mol Biol **53**(3): 401-17.

Sacho, E. J., F. A. Kadyrov, et al. (2008). "Direct visualization of asymmetric adenine-nucleotide-induced conformational changes in MutL alpha." Mol Cell **29**(1): 112-21.

Shibahara, K. and B. Stillman (1999). "Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin." Cell **96**(4): 575-85.

Sogo, J. M., H. Stahl, et al. (1986). "Structure of replicating simian virus 40 minichromosomes. The replication fork, core histone segregation and terminal structures." J Mol Biol **189**(1): 189-204.

Surtees, J. A. and E. Alani (2006). "Mismatch repair factor MSH2-MSH3 binds and alters the conformation of branched DNA structures predicted to form during genetic recombination." J Mol Biol **360**(3): 523-36.

Tran, H. T., J. D. Keen, et al. (1997). "Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants." Mol Cell Biol **17**(5): 2859-65.

Tran, P. T., J. A. Simon, et al. (2001). "Interactions of Exo1p with components of MutLalpha in Saccharomyces cerevisiae." Proc Natl Acad Sci U S A **98**(17): 9760-5.

Uhlmann, F., D. Wernic, et al. (2000). "Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast." Cell **103**(3): 375-86.

von Hippel, P. H. and O. G. Berg (1989). "Facilitated target location in biological systems." J Biol Chem **264**(2): 675-8.

Wagner, G., A. Bancaud, et al. (2005). "Compaction kinetics on single DNAs: purified nucleosome reconstitution systems versus crude extract." Biophys J **89**(5): 3647-59.

Winter, R. B., O. G. Berg, et al. (1981). "Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor--operator interaction: kinetic measurements and conclusions." Biochemistry **20**(24): 6961-77.

Winter, R. B. and P. H. von Hippel (1981). "Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The Escherichia coli repressor--operator interaction: equilibrium measurements." Biochemistry **20**(24): 6948-60.

Chapter 4

A mutation in the putative MLH3 endonuclease domain confers a defect in both mismatch repair and meiosis in *Saccharomyces cerevisiae*: Published and unpublished results

Aaron J. Plys, K.T. Nishant[1] and Eric Alani

Department of Molecular Biology and Genetics
Cornell University
Ithaca, NY 14853-2703

[1]Current address:
Indian Institute of Science Education and Research - Thiruvananthapuram (IISER-TVM), Trivandrum, India

This chapter contains sections taken from the following article originally published in the June 2008 issue of Genetics:

Nishant, K.T., Aaron J. Plys, and Eric Alani. 2008: A mutation in the putative MLH3 endonuclease domain confers a defect in both mismatch repair and meiosis in *Saccharomyces cerevisiae*. Genetics. 2008. Volume 179, Issue 2, Pages 747-755.

Introduction

The mismatch repair (MMR) system is critical for maintenance of genomic integrity. Members of this system are named after *E. coli* factors found to have a "mutator" phenotype when inactivated. Such mutants show an increase in DNA mismatches and insertions/deletion mutations (reviewed in(Modrich 1989). The eukaryotic homologs form defined heterodimeric complexes that show distinct repair properties (Marsischky *et al.* 1996; Wang *et al.* 1999). In *Saccharomyces cerevisiae*, MSH2-MSH6 and MSH2-MSH3 heterodimers are the major components of the mismatch repair pathway that recognize and bind to base-base mismatches and insertion-deletion loops (Johnson *et al.* 1996). MSH-mismatch complexes then recruit a MutL homolog (MLH) heterodimer, which is thought to recruit downstream repair factors (Reviewed in (Jiricny 2006). In mismatch repair the main MLH matchmaker is MLH1-PMS1 (Prolla *et al.* 1994). The MLH1-MLH3 heterodimer has a minor role in the repair of large loops (Flores-Rozas and Kolodner 1998). However, MLH1-MLH3 has an important role in genetic recombination during meiosis. This function is imparted through interactions with the MSH4-MSH5 complex that promotes crossing over between homologous chromosomes (Hunter and Borts 1997).

In most organisms the segregation of homologous chromosomes in the meiosis I (MI) division requires crossing over between homologs (Jones 1987). These crossover events along with sister chromatid cohesion ensure proper tension is generated along the spindle machinery (Maguire 1974). In *S. cerevisiae*, crossover formation is highly regulated such that two crossovers rarely occur within the same genetic interval in the process called positive crossover interference. Crossover interference is thought to ensure that every chromosome,

regardless of size, receives at least one reciprocal exchange event (Jones 1987). In support of this hypothesis is the finding that smaller chromosomes tend to display less positive interference than larger ones (Mortimer and Fogel 1974; Kaback *et al.* 1999; Shinohara *et al.* 2003).

Early in meiosis, programmed double-strand breaks are formed by the Spo11 endonuclease and are later processed into crossovers (Keeney *et al.* 1997). Recent evidence has solidified the model of crossover formation that involves the formation and disappearance of single-end invasion and double-Holliday junction (dHJ) intermediates (Hunter and Kleckner 2001; Bzymek *et al.* 2010). MSH4-MSH5 and MLH1-MLH3 are involved in the major crossover pathway that shows positive interference (de los Santos *et al.* 2003; Argueso *et al.* 2004). One hypothesis is that MSH4-MSH5 stabilizes Holliday junction intermediates (Snowden *et al.* 2004) and MLH1-MLH3 activates and directs an unknown downstream factor that resolves HJ intermediates into crossovers (Hoffmann and Borts 2004; Whitby 2005). The endonuclease MUS81-MMS4 is involved in a separate minor crossover pathway that gives rise to crossovers lacking interference (de los Santos *et al.* 2003). The MUS81-MMS4 endonuclease is thought to create crossovers either by direct resolution of the Holliday junction (HJ) or by cleaving D-loops and half-HJ structures formed in a pre-HJ intermediate (Boddy *et al.* 2001; Hollingsworth and Brill 2004; Gaskell *et al.* 2007). Factors acting in both crossover pathways have been identified in mice and humans, although in contrast to yeast, it appears that crossovers occur primarily through the interference-dependent pathway (Woods *et al.* 1999; Lipkin *et al.* 2002; Santucci-Darmanin *et al.* 2002; Abraham *et al.* 2003; Ciccia *et al.* 2003).

Human MLH1-PMS2 (hMutLα) and yeast MLH1-PMS1 contain an ATP·Mn^{2+}-dependent latent endonuclease activity that is essential for mismatch repair, possibly by providing access to the 5' to 3' exonuclease EXO1 for excision during 3'-directed repair (Kadyrov *et al.* 2006; Kadyrov *et al.* 2007). The $DQHA(X)_2E(X)_4E$ metal-binding motif in the human PMS2 and yeast PMS1 subunits of MutLα was shown to be critical for this activity. Considering that this motif is only found in MutL homologs of organisms lacking MutH has led to the hypothesis that MutL endonuclease activity acts in strand discrimination steps during mismatch repair (Kadyrov *et al.* 2007). Mutations in the metal-binding motif in yeast and humans (*pms2-D699N*, *pms2-E705K*, *pms1-E707K*) abolish both metal co-factor binding and MutLα endonuclease activities, resulting in a defect in mismatch repair (Kadyrov *et al.* 2006; Deschenes *et al.* 2007; Kadyrov *et al.* 2007).

MLH3 homologs in yeast and humans contain the highly conserved metal-binding motif implicated in MutLα endonuclease activity (Kadyrov *et al.* 2006). Dr. K.T. Nishant analyzed point mutations within the endonuclease domain of *S. cerevisiae MLH3* for their effect on meiotic crossing over and the repair of frameshift mutation intermediates (Nishant *et al.* 2008). His genetic analyses illustrated that *mlh3D523N* metal-binding motif point mutants showed phenotypes similar to *mlh3Δ* in both mismatch repair and meiotic crossover assays. These data suggest that the MLH3 endonuclease domain is important for both processes. As mentioned above, genetic and cytological data suggest that MLH3 works at a late step in meiotic recombination while biochemical data on the conserved metal-binding motif in hPMS2 suggest that MLH3 has an endonuclease function. I present here unpublished data on expression and partial purification of yeast MLH1-MLH3. My biochemical data are

inconclusive regarding whether MLH1-MLH3 contains an endonuclease activity *in vitro*. Further studies are necessary to comprehensively address this question.

Results

The *mlh3-D523N* mutation does not affect protein stability or interaction with MLH1. *MLH3* homologs contain a highly conserved DQHA$(X)_2$E$(X)_4$E motif in the C terminus (Figure 4.1) that is thought to be a part of an endonuclease active site (Kadyrov *et al.* 2006). Mutations in this motif disrupted MutLα-dependent (MLH1-PMS1 in yeast, MLH1-PMS2 in humans) endonuclease activities and mismatch repair (Kadyrov *et al.* 2006; Erdeniz *et al.* 2007; Kadyrov *et al.* 2007). To test whether this motif is important for MLH3 functions, we made single and double point mutations in *MLH3* (*D523N, E529K*) that correspond to the human *PMS2* and *S. cerevisiae PMS1* endonuclease mutations (Kadyrov *et al.* 2006; Kadyrov *et al.* 2007). The mutations in *PMS2* and *PMS1* did not affect their protein expression, the stability of the MutLα complex, or interactions between MutLα and MutSα (Kadyrov *et al.* 2006; Kadyrov *et al.* 2007). Predicted *mlh3* endonuclease point mutations were tested for their effect on MLH1-MLH3 interactions using the yeast two-hybrid assay (data not shown and performed by Dr. K.T. Nishant). The interaction between MLH1 and mlh3-D523N was indistinguishable from wild type as measured using *LacZ* and *HIS3* reporters. Two-hybrid interactions between MLH1 and mlh3-E529K and MLH1 and mlh3-D523N, -E529K were not detected. These results are consistent with the *mlh3-D523N* mutation not affecting the stability of MLH3 or the ability of MLH3 to interact with MLH1 and the *mlh3-E529K* mutation disrupting MLH1-MLH3 interactions.

Figure 4.1. Alignment of the conserved DQHA$(X)_2$E$(X)_4$E endonuclease domain between MLH3 homologs.

Location of the conserved endonuclease motif DQHA$(X)_2$E$(X)_4$E in MLH3 homologs in *Mus musculus*(BC079861), *Homo sapiens* (NM_001040108), *S. cerevisiae* (*MLH3*),*Arabidopsis thaliana* (NM_119717) and *Caenorhabditis elegans* (H12C20.2). Alignment was created by K.T. Nishant.

I examined the stability of wild-type and mutant MLH3 in Western blot analysis and during the purification of the MLH1-MLH3 complex. N-terminal HA-tagged versions of MLH3, mlh3-D523N, and mlh3-E529K were created and expressed from *GAL10*, *URA3*, and 2μ plasmids (Materials and Methods). HA-MLH3 and HA-mlh3-E529K were detected at similar levels by Western blot analysis of whole-cell extracts probed with αHA antibody (Figure 4.2A). HA-mlh3-D523N appeared to be expressed at several-fold higher levels. The low signal of strains bearing these overexpression constructs suggests that it would be difficult to detect endogenous levels of HA-MLH3. To test whether mlh3-D523N co-purifies with MLH1, I subjected the supernatant obtained from whole-cell extracts to chitin bead column chromatography. The strains used to make these extracts harbor *MLH3* plasmids (*GALPGK-HA-MLH3*, 2μ, *leu2-d*) that are present at extremely high copy numbers (~400 copies/cell) and plasmids containing *MLH1-VMA1-CBD* fusions (*GAL10-MLH1-VMA1-CBD*, *2*μ, *TRP1*). Because MLH1 is expressed as a chitin-binding domain fusion, this column will retain proteins that co-purify with MLH1 (Hall and Kunkel 2001). MLH1 and associated proteins were then eluted from the column by activating cleavage of the chitin-binding domain from MLH1. Both HA-MLH3 (~83 kDa) and HA-mlh3-D523N were found in the eluate when MLH1 (~87 kDa) was co-overexpressed, whereas HA-mlh3-E529K was not detected. Furthermore, HA-MLH3 was not detected in the absence of MLH1 co-expression (Figure 4.2B).

MLH1-MLH3 is refractory to purification in significant quantities under the conditions tested. Although we can detect HA-MLH3 by Western blot, mass spectrometry analysis of

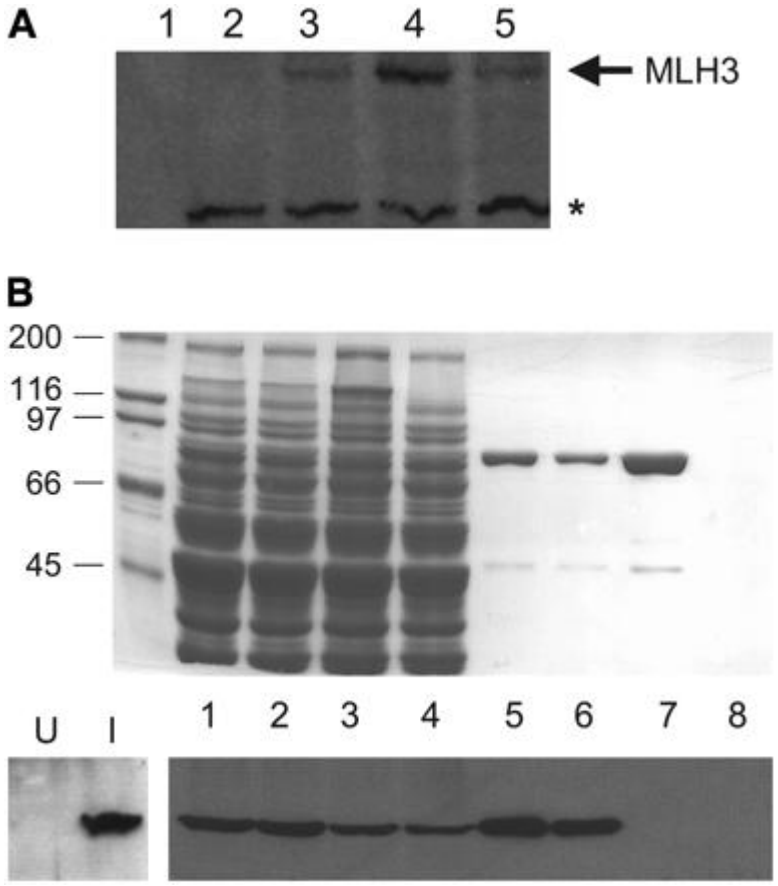
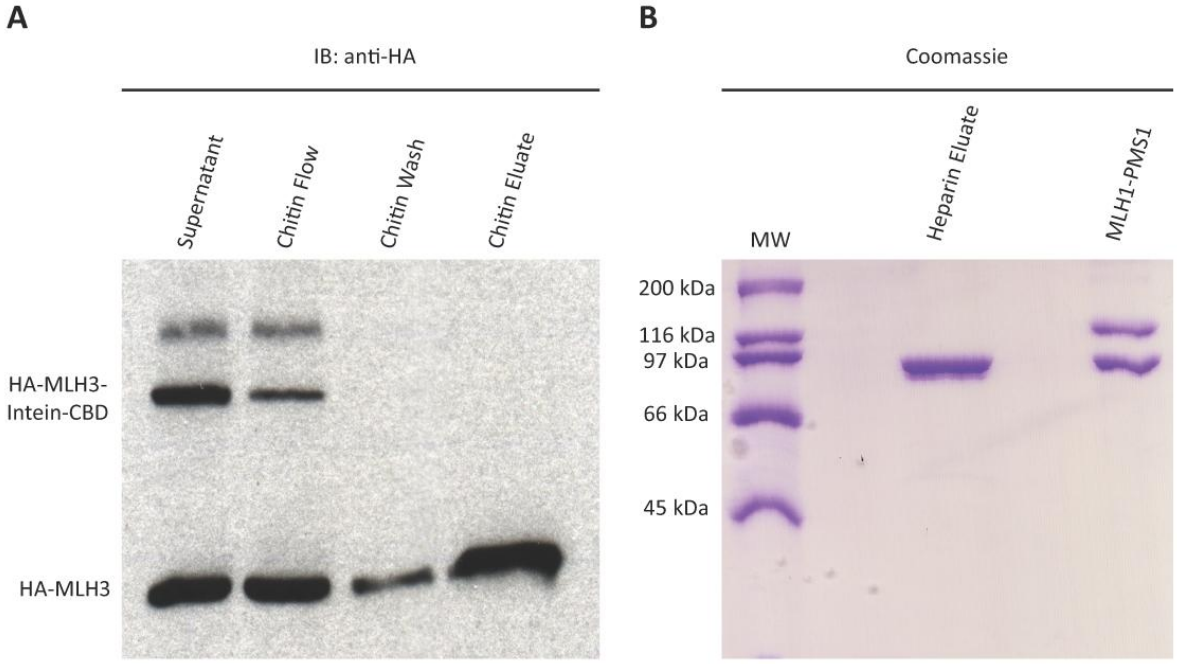


Figure 4.2. Epitope-tagged MLH3 and mlh3-D523N are stably expressed and interact with MLH1.

(A) Crude extracts from galactose-induced yeast containing *GAL10-HA-MLH3*-2µ and *mlh3* derivatives were analyzed in Western blots (8% SDS–PAGE) probed with α-HA antibody (Materials and Methods). Lane 1, purified MLH1-PMS1 (2 µg). Lanes 2–5, cell extracts from strains expressing untagged MLH3 (lane 2), HA-mlh3-E529K (lane 3), HA-mlh3-D523N (lane 4), and HA-MLH3 (lane 5). The asterisk indicates a cross-reacting, nonspecific band.
(B) Partial purification of MLH1-HA-MLH3 and MLH1-HA-mlh3-D523N complexes from immobilized MLH1-VMA1-CBD by chitin bead column chromatography. Eluates from chitin bead columns separated on 8% SDS–PAGE and visualized by Coomassie blue (top) and Western blot analysis with αHA antibody (bottom). Crude extracts are from uninduced (U) or galactose-induced (I) yeast containing *GAL10-HA-MLH3*-2µ-*leu2-d*. Lanes 1–4, input extract (8µl loaded from 50 to 60ml) from cells overexpressing MLH1 and HA-MLH3 (lane 1), MLH1 and HA-mlh3-D523N (lane 2), MLH1 and HA-mlh3-E529K (lane 3), and HA-MLH3 alone (lane 4). Lanes 5–8, pooled chitin bead eluate fractions (8µl loaded from 5.5- to 7.0ml fractions) derived from extracts containing overexpressed MLH1 and HA-MLH3 (lane 5), MLH1 and HA-mlh3-D523N (lane 6), MLH1 and HA-mlh3-E529K (lane 7), and HA-MLH3 alone (lane 8). The sizes of the relevant molecular weight (kDa) standards are indicated.

the band migrating in the Coomassie gel at ~87 kDa revealed MLH1, but not MLH3 peptides (Figure 4.2). The lack of any other Coomassie stained bands suggests that MLH3 is present at very low levels. Low purification yield may be due to inadequate expression and by the indirect nature of purification through immobilized MLH1. In light of this finding, different affinity tag purification procedures were utilized and tested for purification of MLH1-MLH3 complexes. Tags were engineered onto MLH3 to allow for enrichment of this poorly expressed subunit of the complex. Insertion of the chitin binding domain used in the purification in Figure 4.2 onto the C-terminus of *HA-MLH3* (*GALPGK-HA-MLH3-VMA1-CBD, 2*μ, *leu2-d*) followed by co-overexpression with untagged *MLH1* (*GAL10-MLH1, 2*μ, *TRP1*) resulted in detectable expression by Western blot (Figure 4.3A). A high level of intrinsic cleavage of the chitin binding domain off of MLH3 was detected in the supernatant before chitin column chromatography. This results in a loss of over 50% of HA-MLH3 through the inability to be immobilized on the chitin column. Nonetheless, HA-MLH3 was detected in the chitin column eluate and was further purified through an ion-exchange column (Figure 4.3B). Two bands were difficult to distinguish on the Coomassie stained gel, but there appears to be a band of relative higher abundance above another. Since MLH1 is predicted to be larger than MLH3, 87 and 83 kDa, respectively, and MLH1 from purified MLH1-PMS1 appears to co-migrate with the more abundant band, I inferred that the lower band contains MLH3. The identity of each band was confirmed by cutting bands out of the gel and subjecting them individually to mass spectrometry (Figure 4.3C). At present, it is unclear how MLH1 is present at levels above those seen for MLH3 by this purification procedure. It is possible that aggregation may be occurring due to the high levels of MLH1 present, as can be seen during MLH1-PMS1 purification, but these aggregates tend to be refractory to



C

Band	Protein Name	Accession No.	Protein MW	Peptide Count	Protein Score C.I.%	% Protein Coverage
1	DNA mismatch repair protein - MLH1	MLH1_Yeast	87009	43	100	58
2	DNA mismatch repair protein - MLH3	MLH3_Yeast	81948	36	100	45

Figure 4.3. Partial purification of epitope-tagged MLH3 with MLH1.
(A) Partial purification of MLH1-HA-MLH3 complexes by chitin bead column chromatography. All samples are derived from extracts containing overexpressed MLH1 and HA-MLH3-Intein-CBD. Samples were separated on 8% SDS–PAGE and visualized by Western blot analysis with αHA antibody. Uncleaved HA-MLH3-Intein-CBD and cleaved free HA-MLH3 are indicated. **(B)** Purification of chitin eluate by heparin-sepharose ion-exchange chromatography. Samples were separated on 8% SDS–PAGE and visualized by Coomassie staining. Purified MLH1-PMS1 is present for comparison and the sizes of the relevant molecular weight (kDa) standards are indicated. **(C)** Mass spectrometry analysis of bands detected after heparin-sepharose chromatography.

disruption by SDS or boiling. Attempts to remove excess MLH1 monomers resulted in a very low yield where MLH1 was undetectable by Coomassie staining and HA- MLH3 was detected at very low levels by Western blot analysis. Other affinity-tag purification protocols (6xHis and Strept) resulted in similar total yields (data not shown).

MLH1-MLH3 does not contain a detectable endonuclease activity. MLH1-PMS1 was shown to have a latent ATP-Mn^{2+}-dependent endonuclease activity that was detected by nicking of supercoiled plasmid DNA (Kadyrov *et al.* 2006). Partially purified samples from Figure 4.3 were analyzed using a similar nicking assay which resolves uncut supercoiled DNA from nicked products after agarose gel electrophoresis (Figure 4.4). MLH1-MLH3 nicking activity was not detected in the presence of Mn^{2+}, $Mg^{2+,}$ or in combination with ATP (Figure 4.4A). Furthermore, MLH1-MLH3 did not have activity over a broad range of protein concentrations that were well above those required for MLH1-PMS1 activity (Figure 4.4B) (Kadyrov *et al.* 2006). Partially purified MLH1-MLH3 was exhaustively tested under varying conditions of salt, buffer, pH, etc. and was never found to have a reproducible endonuclease activity. My observations are consistent with MLH1-MLH3 not having an endonuclease activity under the purification and conditions analyzed. One caveat to this result is that I was unable to reproduce the endonuclease activity of purified MLH1-PMS1 as characterized by Kadyrov *et al*.

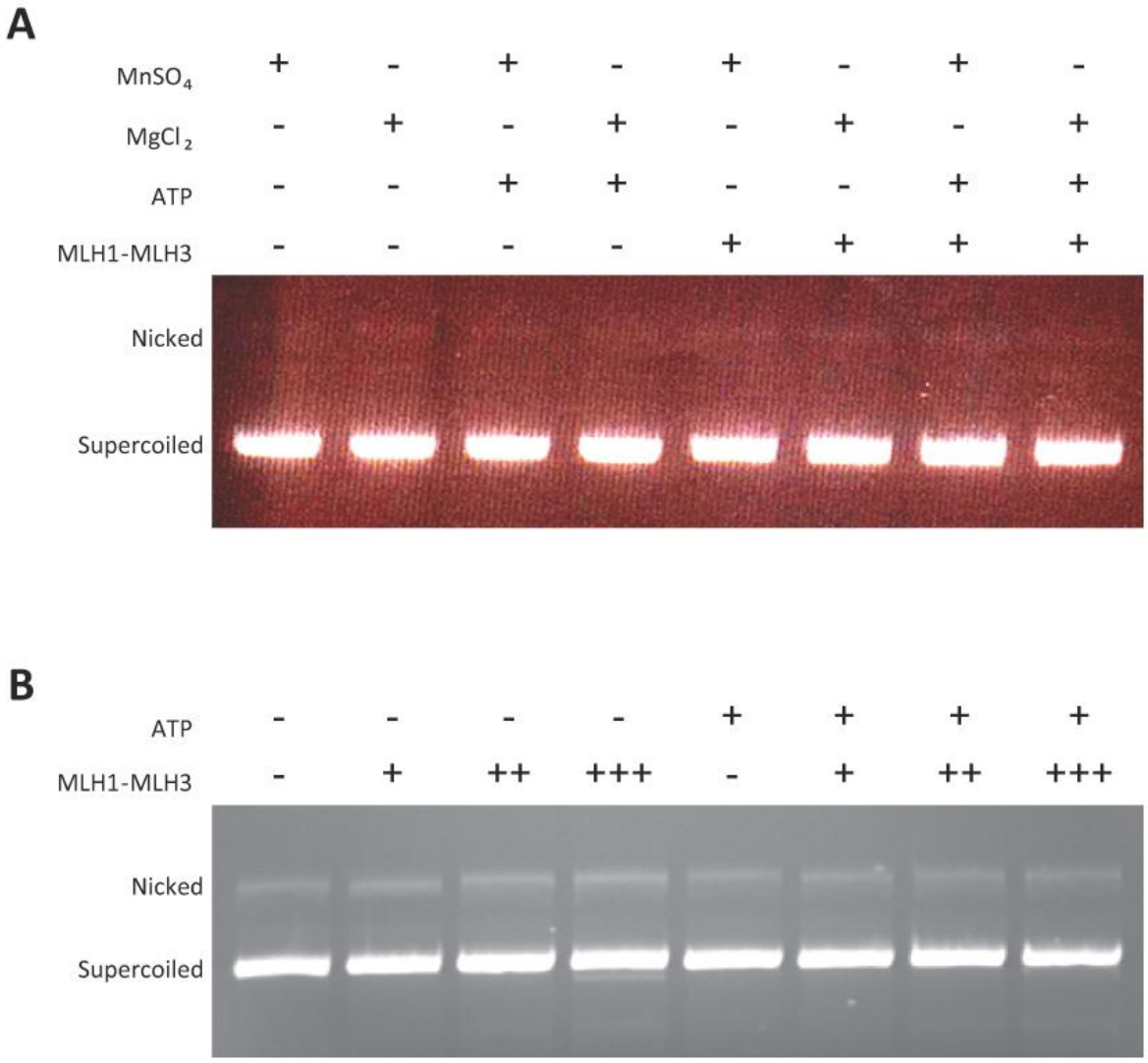


Figure 4.4. Partially purified MLH1-HA-MLH3 does not contain a detectable endonuclease activity.
(A) Supercoiled DNA nicking assay with partially purified MLH1-HA-MLH3. Presence of 1mM $MnSO_4$, 1mM $MgCl_2$, 0.5mM ATP and 80nM MLH1-MLH3 indicate with (+). Supercoiled and nicked DNA indicated. **(B)** Supercoiled DNA nicking assay with increasing amounts of partially purified MLH1-HA-MLH3. All reactions contain 1mM $MnSO_4$. 0.5mM ATP present where indicated. + = 50nM MLH1-MLH3, ++ = 100nM MLH1-MLH3, +++ = 150nM MLH1-MLH3.

Discussion

MLH1-MLH3 is known to play an important role at late stages of interference-dependent crossover recombination. The finding that certain eukaryotic MutL family members, including MLH3, contain a conserved domain associated with endonuclease activity led me to consider if such an activity may facilitate crossover resolution. Genetic analysis by Dr. K.T. Nishant in our lab is consistent with the MLH3 endonuclease domain being critical for both mismatch repair and meiotic recombination functions of MLH1-MLH3. In order to confirm that MLH1-MLH3 possess an endonuclease activity I attempted to purify and biochemically characterize the complex. I have found that the mlh3-D523N endonuclease domain mutant protein appeared to be stable and showed wild-type interactions with MLH1 as measured by column chromatography assays. Only partial purification of the MLH1-MLH3 was obtained under a various purification conditions. Supercoiled DNA nicking assays were inconsistent with these partially purified complexes having any endonuclease activity.

There are several reasons that might explain why I have been unsuccessful at completely purifying the MLH1-MLH3 complex and at showing these complexes have endonuclease activity. It seems unlikely, but there may be a requirement for the last two amino acids of MLH3 for activity since these were removed in order to engineer the HA-MLH3-VMA1-CBD construct. There may also be a detrimental effect of the N-terminal HA epitope tag on activity which can be ruled out by complementation tests in the future. Alternatively, the excess of MLH1 monomers may be inhibitory or the overall yield of MLH1-MLH3 may be too low for utilization in biochemical analysis. Future studies using

alternative expression platforms, such as baculovirus infected insect cells, may circumvent this problem.

MLH1-MLH3 may require additional cofactors such as MSH4-MSH5, MSH2-MSH3 or PCNA for stimulation of latent endonuclease activity. During mismatch repair *in vitro*, the MutLα endonuclease activity was targeted to the discontinuous DNA strand and localized to DNA surrounding the mismatch site. This restriction of MutLα endonuclease activity required mismatch DNA, MutSα, and ATP-Mg^{2+} (Kadyrov *et al.* 2006). On the basis of this information MSH2-MSH3 may be required for MLH1-MLH3 endonuclease activity on loop insertion/deletion mismatches. At present it is not clear in eukaryotes how meiotic crossover products form from Holliday junction intermediates. One possibility is that MLH3 endonuclease activity acts as a Holliday junction resolvase. Alternatively, such an activity could generate precursor recombination intermediates prior to crossover formation. In either model, it may be that MSH4-MSH5 functions to restrict the MLH1-MLH3 endonuclease activity to recombination intermediates that are resolved to crossover products. Such an idea is supported by findings suggesting that MLH1-MLH3 acts downstream of MSH4-MSH5 and that MSH4-MSH5 can bind Holliday junctions *in vitro* (Snowden *et al.* 2004). MLH1-MLH3 may also only be active on relevant substrates such as large loop mismatches and Holliday junction structures. Nonetheless, it is predicted that Mn^{2+} allows more promiscuous cleavage activity for metal-binding nucleases and would be expected to allow MLH1-MLH3 to cleave more types of substrates (Yang *et al.* 2006).

Materials and Methods

***MLH3* plasmids:**

Details on how to make all vectors are available upon request. All mutant alleles were constructed by Quick Change (Stratagene, La Jolla, CA). The mutant alleles were subcloned into wild-type vector backbones and verified by DNA sequencing (Cornell BioResource Center).

pEAE220 (*GAL10-MLH3*, 2µ, *URA3*) contains the *GAL10* promoter driving expression of S288c *MLH3*. pEAE280 (*GAL10-HA-MLH3*, 2µ, *URA3*) is a derivative of pEAE220 that contains an HA-tag (YPYDVPDYA) inserted after the first codon of S288c*MLH3*. pEAE283 (*HA-mlh3-D523N*) and pEAE290 (*HA-mlh3-E529K*) are derivatives of pEAE280 created by Quick Change. These plasmids were used in the Western blots presented in Figure 4.2A. pEAE279 (*GALPGK-HA-MLH3*, 2µ, *leu2-d*) was constructed by inserting the *HA-MLH3* sequence from pEAE280 into the pMMR20 backbone (*GALPGK*, *2*µ, *leu2-d*, a kind gift of L. Prakash). pEAE287 (*GALPGK-HA-mlh3-D523N*, *2*µ, *leu2-d)* and pEAE291 (*GALPGK-HA-mlh3-E529K*, *2*µ, *leu2-d*) are derivatives of pEAE279 that were constructed by Quick Change. These vectors were used for the chitin bead column chromatography presented in Figure 4.2B. pEAE292 (*GALPGK-HA-MLH3-VMA1-CBD, 2*µ, *leu2-d*) was constructed by inserting an HA-MLH3-VMA1-CBD overlap PCR fragment from pEAE279/pMH1 into the pMMR20 backbone. This vector was used for the chitin bead column chromatography presented in Figure 4.3.

Western blots:

BJ2168 (*MAT***a**, *ura3-52*, *trp1-289*, *leu2-3,-112*, *prb1-1122*, *prc1-407*, *pep4-3*) transformed with pMH1(*GAL10-MLH1-VMA1-CBD*, *2*μ, *TRP1*, gift from T. Kunkel) and pEAE220, pEAE280, pEAE283, or pEAE290 was induced in galactose media (0.5-liter inductions yielding ~2 g cell pellets) using methods described by (Hall and Kunkel 2001). Cells were pelleted, washed with chitin buffer (25 mM Tris, pH 8.0, 500 mM NaCl, 10% glycerol, 1 mM EDTA), repelleted, and lysed using glass beads in chitin buffer. SDS–protein loading buffer was added to supernatants after glass bead lysis. Samples were subjected to 8% SDS–PAGE and then transferred to a nitrocellulose membrane. Membranes were blocked with 4% milk overnight and probed with 12CA5 (αHA, Roche) antibody, followed by incubation with α-mouse IgG secondary antibody (Jackson Immunoresearch). Proteins were visualized by the ECL detection method (Amersham/GE).

Protein Purification:

For chitin bead column chromatography with immobilized MLH1, proteins were expressed (5-liter inductions yielding ~17 g of induced cell pellet) from BJ2168 transformed with pMH1 and pEAE279, pEAE287, or pEAE291 as described above. Chitin bead chromatography with immobilized MLH3 was done in the same manner but using strains harboring pEAE49 (*GAL10-MLH1, 2*μ, *TRP1*) and pEAE292. After the chitin-buffer wash step, the cells were resuspended as a thick paste in a small volume of chitin buffer and frozen as drops in liquid nitrogen. The frozen cells were ground up with dry ice in a coffee grinder for lysis. After sublimation of the dry ice overnight at -20°C, cells were thawed on ice and resuspended in 50 ml chitin buffer containing 1 mM PMSF. Chitin bead and Heparin-

sepharose column chromatographies were performed as described (Hall and Kunkel 2001). Briefly, the lysates were cleared by centrifugation and the supernatant was loaded onto a 2 ml chitin bead column equilibrated with chitin buffer. The column was washed with 10 column volumes of chitin buffer followed by 5 column volumes of ATP wash buffer (25 mM Tris, pH 8.0, 200mM KCl, 10% glycerol, 3 mM $MgCl_2$, 1 mM PMSF, 400 µM ATP). Proteins were eluted with 3 column volumes of elution buffer (100 mM Tris, pH 8.0, 100 mM NaCl, 10% glycerol, 1 mM EDTA, 1 mM PMSF, 50 mM DTT). Chitin eluate was loaded onto a 2 ml heparin-sepharose column equilibrated with 0.1 M NaCl Buffer A (25 mM Tris, pH 8.0, 100 mM NaCl, 10% glycerol, 1 mM EDTA, 1 mM PMSF, 10 mM DTT). The column was washed with 5 column volumes of 0.1 M NaCl Buffer A. Proteins were eluted with a 40 ml linear gradient of NaCl (0.1 M-1 M) in buffer A while collecting 1 ml fractions. After SDS–PAGE, protein samples were analyzed by Western blot analysis as described above or stained with Coomassie blue. Yields for purification with MLH1-VMA1-CBD and MLH3 were about 60 µg per prep. Yields for MLH1 with MLH3-VMA1-CBD were about 100 µg per prep. After Heparin-sepharose chromatography of MLH1-MLH3 using the MLH3-VMA1-CBD fusion, peptides were subjected to mass spectrometry. Samples were prepared by cutting bands out of gels after SDS-PAGE and Coomassie blue staining. Mass spectrometry analysis was conducted by Cornell University Life Science Core Laboratories Center.

Endonuclease assay:

Supercoiled DNA nicking assays were performed as described in (Kadyrov *et al.* 2006). Briefly, 40 µl reactions containing 23 mM Hepes-KOH, pH 7.4, 0.5 mg/ml BSA, 23 mM NaCl, 1 mM DTT, 2% glycerol, 0.2 µg of supercoiled DNA (pEAE220 purified by CsCl/EtBr

gradients (Sambrook *et al.* 1989)) and partially purified MLH1-HA-MLH3 were incubated at 30˚C for 20 minutes. Reactions were terminated by the addition of 0.1% SDS, 14 mM EDTA and 0.1 mg/ml Proteinase K. After further incubation at 55°C for 15 min, PMSF was added to 4 mM, and products were resolved by electrophoresis through 0.8% agarose and stained with ethidium bromide.

Acknowledgements

I would like to thank members of the Alani laboratory with a special thank you to K.T. Nishant for his help on this project. Also I would like to thank Paula Cohen, Neil Hunter, and Tom Kunkel for thoughtful discussions and comments on the manuscript that was included in this chapter. This work was supported by National Institutes of Health grant GM53085 (E.A. and K.T.N.) and a State University of New York fellowship (A.J.P.).

References

Abraham, J., B. Lemmers, et al. (2003). "Eme1 is involved in DNA damage processing and maintenance of genomic stability in mammalian cells." Embo J **22**(22): 6137-47.

Argueso, J. L., J. Wanat, et al. (2004). "Competing crossover pathways act during meiosis in Saccharomyces cerevisiae." Genetics **168**(4): 1805-16.

Boddy, M. N., P. H. Gaillard, et al. (2001). "Mus81-Eme1 are essential components of a Holliday junction resolvase." Cell **107**(4): 537-48.

Bzymek, M., N. H. Thayer, et al. (2010). "Double Holliday junctions are intermediates of DNA break repair." Nature **464**(7290): 937-41.

Ciccia, A., A. Constantinou, et al. (2003). "Identification and characterization of the human mus81-eme1 endonuclease." J Biol Chem **278**(27): 25172-8.

de los Santos, T., N. Hunter, et al. (2003). "The Mus81/Mms4 endonuclease acts independently of double-Holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast." Genetics **164**(1): 81-94.

Deschenes, S. M., G. Tomer, et al. (2007). "The E705K mutation in hPMS2 exerts recessive, not dominant, effects on mismatch repair." Cancer Lett **249**(2): 148-56.

Erdeniz, N., M. Nguyen, et al. (2007). "Mutations affecting a putative MutLalpha endonuclease motif impact multiple mismatch repair functions." DNA Repair (Amst) **6**(10): 1463-70.

Flores-Rozas, H. and R. D. Kolodner (1998). "The Saccharomyces cerevisiae MLH3 gene functions in MSH3-dependent suppression of frameshift mutations." Proc Natl Acad Sci U S A **95**(21): 12404-9.

Gaskell, L. J., F. Osman, et al. (2007). "Mus81 cleavage of Holliday junctions: a failsafe for processing meiotic recombination intermediates?" Embo J **26**(7): 1891-901.

Hall, M. C. and T. A. Kunkel (2001). "Purification of eukaryotic MutL homologs from Saccharomyces cerevisiae using self-affinity technology." Protein Expr Purif **21**(2): 333-42.

Hoffmann, E. R. and R. H. Borts (2004). "Meiotic recombination intermediates and mismatch repair proteins." Cytogenet Genome Res **107**(3-4): 232-48.

Hollingsworth, N. M. and S. J. Brill (2004). "The Mus81 solution to resolution: generating meiotic crossovers without Holliday junctions." Genes Dev **18**(2): 117-25.

Hunter, N. and R. H. Borts (1997). "Mlh1 is unique among mismatch repair proteins in its ability to promote crossing-over during meiosis." Genes Dev **11**(12): 1573-82.

Hunter, N. and N. Kleckner (2001). "The single-end invasion: an asymmetric intermediate at the double-strand break to double-holliday junction transition of meiotic recombination." Cell **106**(1): 59-70.

Jiricny, J. (2006). "The multifaceted mismatch-repair system." Nat Rev Mol Cell Biol **7**(5): 335-46.

Johnson, R. E., G. K. Kovvali, et al. (1996). "Requirement of the yeast MSH3 and MSH6 genes for MSH2-dependent genomic stability." J Biol Chem **271**(13): 7285-8.

Jones, G. H. (1987). Chiasmata. Meiosis. P. Moens. New York/London/San Diego, Academic Press**:** 213-238.

Kaback, D. B., D. Barber, et al. (1999). "Chromosome size-dependent control of meiotic reciprocal recombination in Saccharomyces cerevisiae: the role of crossover interference." Genetics **152**(4): 1475-86.

Kadyrov, F. A., L. Dzantiev, et al. (2006). "Endonucleolytic function of MutLalpha in human mismatch repair." Cell **126**(2): 297-308.

Kadyrov, F. A., S. F. Holmes, et al. (2007). "Saccharomyces cerevisiae MutLalpha is a mismatch repair endonuclease." J Biol Chem **282**(51): 37181-90.

Keeney, S., C. N. Giroux, et al. (1997). "Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family." Cell **88**(3): 375-84.

Lipkin, S. M., P. B. Moens, et al. (2002). "Meiotic arrest and aneuploidy in MLH3-deficient mice." Nat Genet **31**(4): 385-90.

Maguire, M. P. (1974). "Letter: The need for a chiasma binder." J Theor Biol **48**(2): 485-7.

Marsischky, G. T., N. Filosi, et al. (1996). "Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair." Genes Dev **10**(4): 407-20.

Modrich, P. (1989). "Methyl-directed DNA mismatch correction." J Biol Chem **264**(12): 6597-600.

Mortimer, R. and S. Fogel (1974). Genetical interference and gene conversion. Mechanisms in Recombination. R. Grell. New York, Plenum Press**:** 263-275.

Nishant, K. T., A. J. Plys, et al. (2008). "A mutation in the putative MLH3 endonuclease domain confers a defect in both mismatch repair and meiosis in Saccharomyces cerevisiae." Genetics **179**(2): 747-55.

Prolla, T. A., Q. Pang, et al. (1994). "MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast." Science **265**(5175): 1091-3.

Sambrook, J., F. E.F., et al. (1989). Molecular Cloning. Cold Spring Harbor, Cold Spring Harbor Laboratory Press.

Santucci-Darmanin, S., S. Neyton, et al. (2002). "The DNA mismatch-repair MLH3 protein interacts with MSH4 in meiotic cells, supporting a role for this MutL homolog in mammalian meiotic recombination." Hum Mol Genet **11**(15): 1697-706.

Shinohara, M., K. Sakai, et al. (2003). "Crossover interference in Saccharomyces cerevisiae requires a TID1/RDH54- and DMC1-dependent pathway." Genetics **163**(4): 1273-86.

Snowden, T., S. Acharya, et al. (2004). "hMSH4-hMSH5 recognizes Holliday Junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes." Mol Cell **15**(3): 437-51.

Wang, T. F., N. Kleckner, et al. (1999). "Functional specificity of MutL homologs in yeast: evidence for three Mlh1-based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction." Proc Natl Acad Sci U S A **96**(24): 13914-9.

Whitby, M. C. (2005). "Making crossovers during meiosis." Biochem Soc Trans 33(Pt 6): 1451-5.

Woods, L. M., C. A. Hodges, et al. (1999). "Chromosomal influence on meiotic spindle assembly: abnormal meiosis I in female Mlh1 mutant mice." J Cell Biol 145(7): 1395-406.

Yang, W., J. Y. Lee, et al. (2006). "Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity." Mol Cell 22(1): 5-13.

Chapter 5

Implications for the DNA mismatch repair field and unresolved questions

Implications for the diffusion properties of the DNA mismatch recognition complexes. My work has provided important insights into the mechanism of movement utilized by MLH1-PMS1 to scan the genome in search of targets. MLH1-PMS1 diffuses very rapidly on DNA via a hopping/stepping mechanism (Gorman *et al.* 2010); Chapter 2). This type of diffusion mechanism allows MLH1-PMS1 complexes to efficiently bypass obstacles on DNA such as nucleosomes. This was in striking contrast to the mismatch recognition complex MSH2-MSH6, which uses a sliding diffusion mechanism and is unable to traverse nucleosome barriers (Gorman *et al.* 2010); Chapter2). In Chapter 3, I presented a model for the diffusion properties of each mismatch repair complex in the context of an active replication fork (Figure 3.1). In this model MSH2-MSH6 tracks with the replication fork proteins to rapidly identify mismatches as they are created. Importantly, this is in an environment that is thought to be transiently devoid of nucleosomes that would otherwise be potential barriers to mismatch recognition by MSH2-MSH6. The ability of MLH1-PMS1 to traverse nucleosomes would allow for efficient detection of MSH2-MSH6 bound to a mismatch after nucleosomes are reestablished downstream of the fork.

I found that the DNA binding properties of MLH1-PMS1 appear to be modulated by unstructured linker arms in each subunit that connect to well-defined terminal globular domains. In particular, the length of the linker arm is important for function *in vivo*. DNA

binding and MSH2-MSH6 interactions remain intact with purified complexes containing shortened linker arms, indicating that these complexes are defective in some other aspect of MLH1-PMS1 function. As mentioned in Chapter 3, future single-molecule work will be able to confirm my hypothesis that the lengths of the linker arms of MLH1-PMS1 are important for the rapid diffusion along DNA and bypass of nucleosomes that was presented in Chapter 2.

Another important question to address is the behavior of both MSH2-MSH6 and MLH1-PMS1 after ternary complex formation. It is unclear how strand discrimination is coordinated with mismatch recognition but it is thought to involve protein movement along the DNA (Gradia *et al.* 1997; Gradia *et al.* 1999; Pluciennik and Modrich 2007). It seems likely that the mismatch repair complexes would be able to use one-dimensional diffusion to search out the unknown strand discrimination signal in a manner analogous to their initial search for mismatches. Indeed, studies have shown that MSH2-MSH6 can undergo a nucleotide-dependent transformation into a hydrolysis-independent sliding clamp that disengages from the mismatch site, but these studies do not include the behavior of MLH1-PMS1 (Gradia *et al.* 1997; Gradia *et al.* 1999). Successive loading and disengagement events could allow for polymer tract formation connecting the mismatch to the strand signal (Figure 5.1). In support of this is the finding that MLH1-PMS1 displays cooperative binding and an ability to form continuous tracts on duplex DNA (Hall *et al.* 2001). Furthermore, the environment encountered during the strand discrimination search would likely be a chromatin landscape that would serve as a barrier to MSH2-MSH6 diffusion (Li *et al.* 2009; Gorman *et al.* 2010). Intriguingly, hMutsα (MSH2-MSH6) has been found to suppress histone deposition by CAF-I in an *in vitro* assay that requires a mismatch (Kadyrova *et al.* 2010). This suggests a

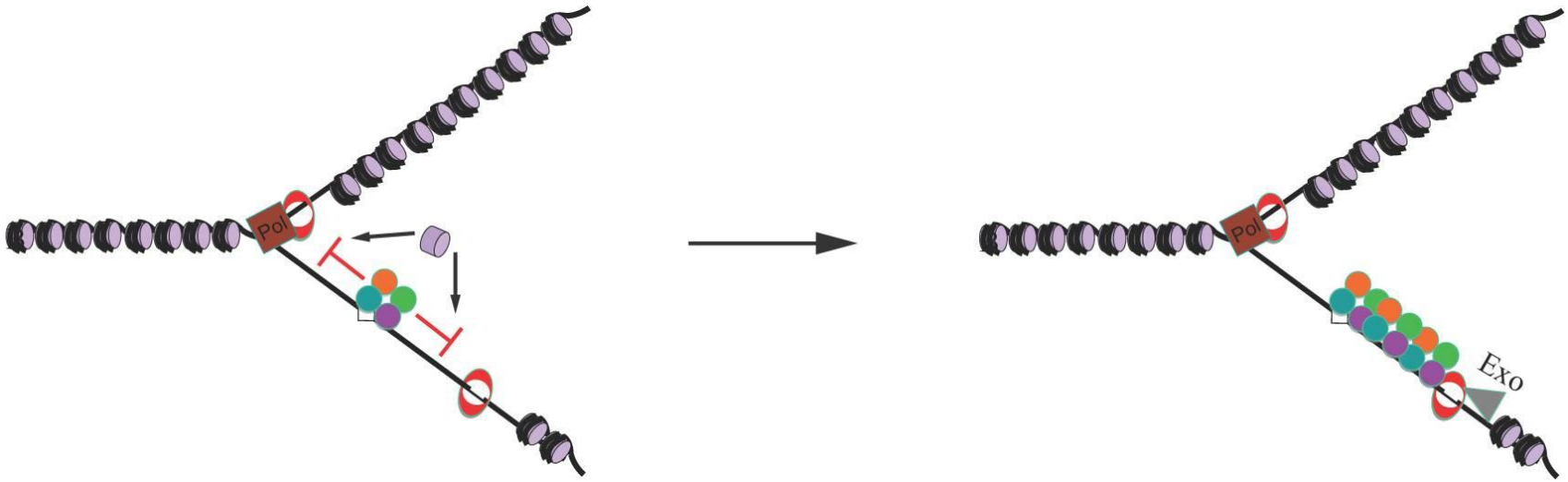


Figure 5.1. Speculative model for strand discrimination signal search by DNA mismatch repair factors after mismatch recognition. The left panel illustrates a ternary complex of MSH2-MSH6 (blue and purple circles) and MLH1-PMS1 (orange and green circles) at a mismatch (diamond) on the lagging strand of a model replication fork. Local inhibition of nucleosome deposition around the mismatch site by the mismatch repair factors may create an area free of obstacles in order to efficiently search for the unknown strand discrimination signal by 1D diffusion. The right panel illustrates 1D diffusion of mismatch repair complexes towards potential strand discrimination signals. The strand discrimination signal has been proposed to be PCNA (Red ring), a strand discontinuity (white gap) or both (Jiricny 2006; Modrich 2006). Multiple mismatch repair complexes loading and disengaging from the mismatch could lead to a polymer formation connecting the mismatch to the strand discrimination signal. Subsequently, EXO1 is loaded and activated on the newly replicated strand to degrade nucleotides passed the mismatch to create a gap that is filled in by DNA polymerase and DNA ligase.

local inhibition of replication-dependent nucleosome assembly when a mismatch is encountered in order to allow efficient repair events (Figure 5.1). This also fits nicely with our model of mismatch recognition happening in concert with the replication fork machinery in order to quickly correct errors before permanent barriers are created. Another potential mechanism to remove inhibitory nucleosomes that has not been clearly established is the recruitment or activation of chromatin remodeling activities to regions surrounding mismatches (Gavin *et al.* 2006; Javaid *et al.* 2009).

The most significant unanswered question in the eukaryotic DNA mismatch repair field is the nature of the strand discrimination signal. Single molecule studies have the potential to address this enigma following the characterization of MSH2-MSH6 and MLH1-PMS1 behaviors after mismatch binding. In this system it will be interesting to test how the proposed strand signals (single-stranded nicks or PCNA) influence the diffusion behaviors of the mismatch repair complexes (Jiricny 2006; Modrich 2006). Furthermore, the minimal eukaryotic DNA mismatch repair system has been reconstituted in an *in vitro* assay that gives hope for the ultimate goal of visualizing the repair reaction through single molecule techniques (Genschel and Modrich 2003; Dzantiev *et al.* 2004; Zhang *et al.* 2005).

Mechanistic details of the DNA binding activity of MLH1-PMS1. Structural studies have led to the proposal that MLH1-PMS1 could adopt a ring like structure capable of encircling DNA (Guarne *et al.* 2004). TIRFM experiments are consistent with MLH1-PMS1 topologically binding DNA and preferentially dissociated from free DNA ends (Gorman *et al.* 2010). In support of this, TEV cleavage of the unstructured linker arms of MLH1-PMS1 results in a loss of DNA binding presumably through opening of the protein ring (Gorman *et*

al. 2010). MLH1 can form a homodimeric complex in the absence of PMS1 that is non-functional in mismatch repair (Shcherbakova *et al.* 2001). Unlike the heterodimeric complex, the N-terminal domains of MLH1 homodimers do not associate and thus the complex does not form a ring (Hall *et al.* 2003). Surprisingly, MLH1 homodimers bind DNA and diffuse at a 7-fold faster rate than MLH1-PMS1. However, homodimeric MLH1 does not show end-dependent dissociation, indicating that PMS1 is necessary for topological DNA binding (Gorman *et al.* 2010). These results led me to wonder if the N-terminal domains of MLH1 and PMS1 could act independently of their corresponding C-terminal domains.

Here I present preliminary data and future directions that are meant to test this hypothesis. I created swap alleles of MLH1 and PMS1 that have the N-terminal domain of one subunit attached to the unstructured linker arm and C-terminal fragment of the other (See Appendix, Supplementary Materials and Methods). These constructs give rise to three potential complexes (Appendix, Supplementary Figure 5.2): 1. Swap allele containing the N-terminus of MLH1 connected to the C-terminus of PMS1 plus full-length MLH1. This complex could dimerize through C-terminal associations similar to the heterodimer, but would contain only MLH1 N-termini that should not associate. Therefore, this complex should mimic the characteristics of MLH1 homodimers. 2. Swap allele containing the N-terminus of PMS1 connected to the C-terminus of MLH1 plus full-length MLH1. This complex could dimerize through C-terminal associations similar to MLH1 homodimers, but would contain the heterodimeric pair of N-termini. If the N-terminal domains can function independently of one another then this complex should behave like heterodimeric MLH1-PMS1. 3. Both swap alleles together. This complex contains the heterodimeric pair of both N-

and C-termini but in swapped configuration. I would expect that if the N-termini can function independently, then this complex should also behave like wild-type MLH1-PMS1.

Using the *lys2*::*insE-A*$_{14}$ mismatch repair assay I found that all combination of swap alleles displayed an increase in mutation rates similar to *mlh1Δ pms1Δ*. (Appendix, Supplementary Table 5.1). These results indicate that swapping N-terminal domains is not possible for functional MLH1-PMS1 complexes. In the future the expression and stability of these swap constructs need to be verified to confirm these results. It might also be useful to see if purified complexes can bind DNA determine their diffusion patterns.

Another potential method to test if MLH1-PMS1 encircles DNA, as well as, the importance of periodic ring opening is through chemically induced heterodimerization. In this method, chimeric proteins containing the 12 kDa FK506-binding protein (FKBP12) and the 89 amino acid FKBP12-rapamycin-binding (FRB) domain are induced to dimerize by the drug rapamycin (Xu *et al.* 2010). MLH1 and PMS1 are known to dimerize between their C-terminal globular domains and it has been suggested that interactions between their N-terminal globular domains may be modulated by nucleotide binding (Tran and Liskay 2000; Guarne *et al.* 2004). Engineering chimeric MLH1 and PMS1 proteins that contain the FKBP12 and FRB sequences in their N-terminal globular domains will allow for chemically inducible ring formation that can be experimentally tested for both *in vitro* and *in vivo* function. This analysis will be able to determine if MLH1-PMS1 needs to be able to open from a ring-like structure to initially gain access to DNA and can also test if proteins that are bound to DNA display altered diffusion properties when induced to dimerize. I would expect that chimeric MLH1-PMS1 would display decreased DNA binding activity if treated with rapamycin before interacting with DNA due to an inability of the DNA to enter the enclosed

protein ring. If MLH1-PMS1 does indeed encircle the DNA as it diffuses, then proteins that are pre-bound to DNA and then treated with rapamycin would be expected to show similar properties to complexes that were untreated. This analysis may also be extended to MLH1 homodimers in order to test if chemically induced dimerization results in DNA diffusion characteristics more similar to MLH1-PMS1 heterodimers.

References

Dzantiev, L., N. Constantin, et al. (2004). "A defined human system that supports bidirectional mismatch-provoked excision." Mol Cell **15**(1): 31-41.

Gavin, A. C., P. Aloy, et al. (2006). "Proteome survey reveals modularity of the yeast cell machinery." Nature **440**(7084): 631-6.

Genschel, J. and P. Modrich (2003). "Mechanism of 5'-directed excision in human mismatch repair." Mol Cell **12**(5): 1077-86.

Gorman, J., A. J. Plys, et al. (2010). "Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice." Nat Struct Mol Biol **17**(8): 932-8.

Gradia, S., S. Acharya, et al. (1997). "The human mismatch recognition complex hMSH2-hMSH6 functions as a novel molecular switch." Cell **91**(7): 995-1005.

Gradia, S., D. Subramanian, et al. (1999). "hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA." Mol Cell **3**(2): 255-61.

Guarne, A., S. Ramon-Maiques, et al. (2004). "Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair." Embo J **23**(21): 4134-45.

Hall, M. C., P. V. Shcherbakova, et al. (2003). "DNA binding by yeast Mlh1 and Pms1: implications for DNA mismatch repair." Nucleic Acids Res **31**(8): 2025-34.

Hall, M. C., H. Wang, et al. (2001). "High affinity cooperative DNA binding by the yeast Mlh1-Pms1 heterodimer." J Mol Biol **312**(4): 637-47.

Javaid, S., M. Manohar, et al. (2009). "Nucleosome remodeling by hMSH2-hMSH6." Mol Cell **36**(6): 1086-94.

Jiricny, J. (2006). "The multifaceted mismatch-repair system." Nat Rev Mol Cell Biol **7**(5): 335-46.

Kadyrova, L. Y., E. R. Blanko, et al. (2010). "CAF-I-dependent control of degradation of the discontinuous strands during mismatch repair." Proc Natl Acad Sci U S A **108**(7): 2753-8.

Li, F., L. Tian, et al. (2009). "Evidence that nucleosomes inhibit mismatch repair in eukaryotic cells." J Biol Chem **284**(48): 33056-61.

Modrich, P. (2006). "Mechanisms in eukaryotic mismatch repair." J Biol Chem **281**(41): 30305-9.

Pluciennik, A. and P. Modrich (2007). "Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair." Proc Natl Acad Sci U S A **104**(31): 12709-13.

Shcherbakova, P. V., M. C. Hall, et al. (2001). "Inactivation of DNA mismatch repair by increased expression of yeast MLH1." Mol Cell Biol **21**(3): 940-51.

Tran, P. T. and R. M. Liskay (2000). "Functional studies on the candidate ATPase domains of Saccharomyces cerevisiae MutLalpha." Mol Cell Biol **20**(17): 6390-8.

Xu, T., C. A. Johnson, et al. (2010). "Conditionally controlling nuclear trafficking in yeast by chemical-induced protein dimerization." Nat Protoc **5**(11): 1831-43.

Zhang, Y., F. Yuan, et al. (2005). "Reconstitution of 5'-directed human mismatch repair in a purified system." Cell **122**(5): 693-705.

Appendix

Chapter 2 Supplemental Discussion 152

Chapter 2 Supplementary Figures 154

Chapter 5 Supplementary Materials and Methods 159

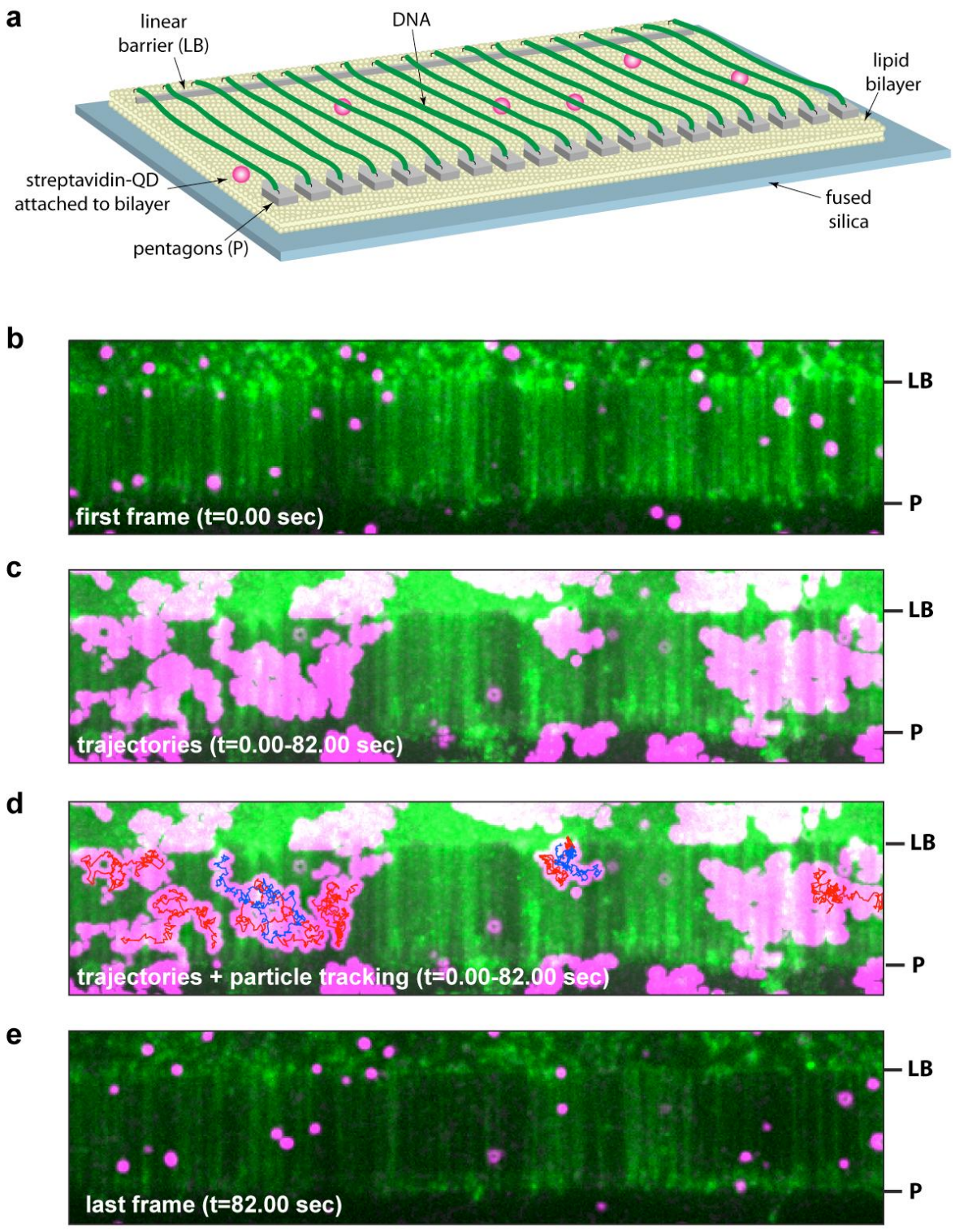
Chapter 5 Supplementary Figure and Table 160

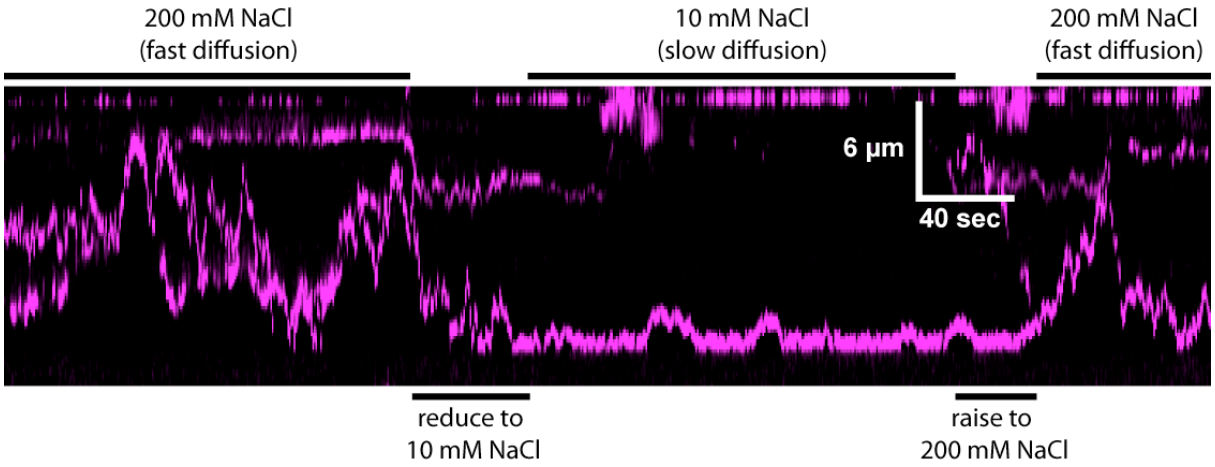
Chapter 2 Supplemental Discussion

Mechanisms of facilitated diffusion and effects of salt. Proteins that hop cycle rapidly between a free and a bound state, and increasing ionic strength increases the lifetime of unbound intermediate while decreasing the lifetime of the bound intermediate, hence increasing the overall observed rate of travel (Berg *et al.*, Biochemistry, 1981; Blainey et al., PNAS, 2006; Kochaniak, *et al*., J. Biol. Chem., 2008; Komazin-Meredith, *et al*., PNAS, 2008). Stepping can be considered virtually identical to hopping, with the exception that at least two separate parts of the protein must cycle between free and bound states (*i.e.* hop). The effect of salt on a stepping process would be similar to that which is observed for a simpler hopping mechanism and would lead to an apparent overall increase in the rate of travel along the DNA. It is important to note that different forms of facilitated diffusion are not mutually exclusive, and bound states that exist during a hopping/stepping mechanism may in fact slide on DNA (Givaty & Levy, J. Mol. Biol., 2009). Jumping is distinct from hopping/stepping, in that it is an uncorrelated search involving a free 3D diffusion component enabling the protein to move long distances between each independent jumping event (von Hippel & Berg, J. Biol. Chem. 1989). Similar to hopping/stepping, jumping frequency would also increase at higher ionic strengths, which in the case of jumping would lead to increased dissociation of the protein from the DNA, and single molecule experiments done in the presence of buffer flow induce dissociation of jumping proteins, as the free state is readily pushed away from the DNA and irretrievably lost to solution. We do not completely rule out the possibility that occasional jumping could contribute to Mlh1-Pms1 movement. However, in a single molecule assay, jumping would appear as the sudden disappearance of a protein followed by its near

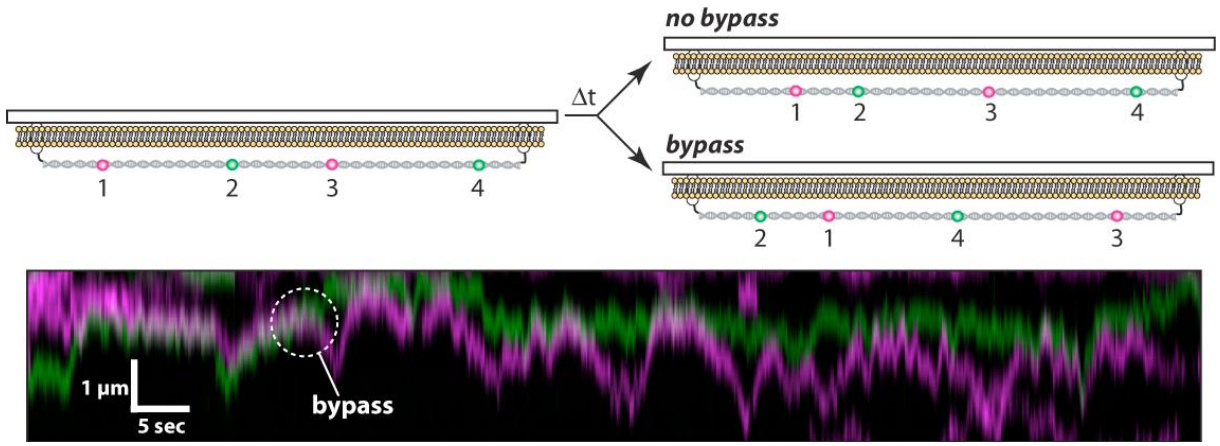
immediate reappearance at a distant location or even on a different DNA molecule (Bonnet *et al.*, Nucleic Acids Res. 2008). The vast majority of the diffusion trajectories observed for Mlh1-Pms1 involved continuous 1D motion along the DNA, which is only consistent with a correlated scanning mechanism (*i.e.* sliding and/or hopping/stepping), but inconsistent with extensive jumping. Moreover, extensive jumping is inconsistent with the end-dependent DNA dissociation observed for Mlh1-Pms1 and the wrapped DNA-binding topology that we propose as an explanation for the end-dependent dissociation.

Supplementary Figure 2.7. DNA curtains provide sufficient clearance for free passage of QDs. (A) Schematic overview of the experimental setup. A double-tethered DNA curtain (green) was prepared as described in the Materials and Methods. Streptavidin-QDs (Invitrogen; Magenta) were then anchored to the lipid bilayer, which contains a subset of biotinylated lipids, and videos were collected to determine if the anchored QDs could diffuse underneath the DNA. **(B)** Image of the DNA curtain and QDs at the start of the experiment. **(C)** Image showing numerous QD trajectories over an 82 second period and **(D)** eight examples of 2D particle tracking (blue or red traces) detailing movement of individual QDs as they diffuse underneath the DNA molecules. **(E)** Final image from the data set. These results demonstrate that the anchored DNA molecules are far enough away from the surface of the microfluidic sample chamber to allow unhindered passage of a QD.


a
linear
barrier (LB)
DNA
lipid
bilayer
streptavidin-QD
attached to bilayer
fused
silica
pentagons (P)
b
LB
P
first frame (t=0.00 sec)
c
LB
P
trajectories (t=0.00-82.00 sec)
d
LB
P
trajectories + particle tracking (t=0.00-82.00 sec)
e
LB
P
last frame (t=82.00 sec)



Supplementary Figure 2.8. Salt-dependence of Mlh1-Pms1 1D diffusion. Mlh1-Pms1 was bound to a double-tethered DNA molecule at 200 mM NaCl and allowed to diffuse in the absence of buffer flow. At the 160-second time point the flow chamber was gently washed with buffer containing 10 mM NaCl (as indicated). The wash was completed at the 232-second time point and Mlh1-Pms1 was monitored in the absence of buffer flow. A second wash was initiated at 360-seconds to raise the NaCl concentration back up to 200 mM, and this was completed at 420-seconds. Buffer flow was then terminated and diffusion was allowed to proceed in the absence of flow. As shown here, the movement of Mlh1-Pms1 was highly dependent upon the concentration of NaCl, and the protein complex rapidly diffuses at high salt (D1d=0.310 µm2 sec-1 at 200 mM NaCl), but diffuses much more slowly when the salt was reduced (D1d=0.021 µm2 sec-1 at 10 mM NaCl; corresponding to a 14-fold decrease in the diffusion coefficient), and then begins rapid diffusion once the salt concentration was increased again (D1d=0.281 µm2 sec-1 at 200 mM NaCl).



Supplementary Figure 2.9. Mlh1-Pms1 complexes can occasionally bypass one another while traveling along the same DNA. Two Mlh1-Pms1 molecules are bound to the same DNA and each is labeled with a different colored QD (either green or magenta). An example of a bypass event is highlighted. We observed a total of 5 bypass events out of ~50 pairs of colliding magenta and green Mlh1-Pms1 complexes. Please note that each of these two-color pairs collided with one another numerous times, but we cannot determine the exact number of collisions because we currently lack the spatial resolution to know whether the proteins are in actual physical contact with one another or are just very close together.

Supplementary Table 2.1. Summary of Mlh1-Pms1 behavior from videos.

Stuck to barrier	***127*** †
Stuck to bilayer	***14*** †
Bound to DNA	***203***
Stationary on DNA	**32**
stationary, not blinking	*(25)* †
stationary, blinking	*(7)*
Moving on DNA	**172** §
not blinking	*(27)* †
too many collisions	*(75)*
lifetime too short	*(48)*
trackable	*(22)* ‡
Total QDs	***344***

† Excluded from further consideration.

‡ Trackable complexes over 250 consectutive frames and did not collide with other proteins.

§ The percent of DNA-bound complexes that are moving is equal to the number of DNA bound complexes that were moving (145) divided by the total number of DNA -bound complexes (152) and yields a value of 95.4%. This calculation excludes any QDs that were not blinking.

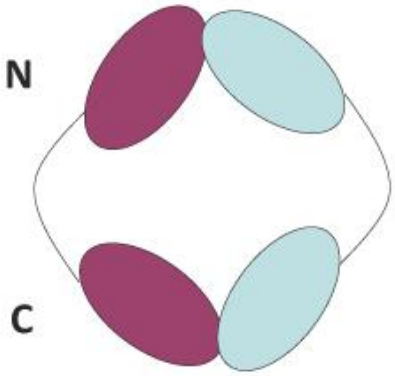
Chapter 5 Supplementary Materials and Methods

Construction of MLH1-PMS1 swap alleles. *MLH1* and *PMS1* N-terminal domain swaps were constructed by overlap extension PCR. Plasmids containing the native promoter associated with the resultant N-terminal domain after swap construction were used to drive expression of each allele in complementation assays. pEAA551 is a derivative of pEAA213 (see Table 3.1) that contains a swap allele that fuses the NTD of *MLH1* with the linker arm and CTD of *PMS1* under the native *MLH1* promoter. The fusion point is after amino acid 335S of MLH1 and before amino acid 390T of PMS1. pEAA552 is a derivative of pEAA238 (See Table 3.1) that contains a swap allele that fuses the NTD of *PMS1* with the linker arm and CTD of *MLH1* under the native *PMS1* promoter. The fusion point is after amino acid 389T of PMS1 and before amino acid 336A of MLH1All clones were sequenced (Cornell BioResource Center), and additional details on vector construction will be provided upon request.

All swap alleles were tested *in vivo* for complementation of the *mlh1Δ pms1Δ* mutator phenotype in EAY 1365 using the *lys2*::*insE-A*$_{14}$ reversion assay as described in Chapter 3 Materials and Methods. All complementation assays were carried out in EAY1365 (see Table 3.1) with the indicated allele combinations (See Supplementary Table 5.1). pEAA213 was present in combinations that included the wild-type *MLH1* allele.

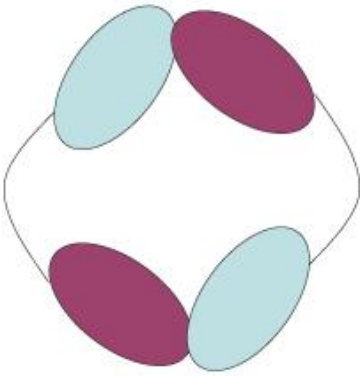
Supplementary Figure 5.2. Illustration of complexes containing engineered N-terminal swaps of MLH1-PMS1. (A) Illustration of wild-type MLH1-PMS1 showing the complex in a ring configuration. The N-termini are at the top of the complex indicated by an **N** and the C-termini are at the bottom indicated by a **C**. Lines connecting the two globular domains represent the unstructured linker arms **(B)** Complex containing a swap of both N-terminal domains of the MLH1-PMS1 complex. Each subunit was engineered by overlap extension PCR to replace the existing N-terminal domain including the unstructured linker arm with the same region from the other subunit. **(C)** Complex containing a swap of the N-terminal domain of PMS1 with that of MLH1 in a complex with full-length MLH1. **(D)** Complex containing a swap of the N-terminal domain of MLH1 with that of PMS1 in a complex with full-length MLH1.

A



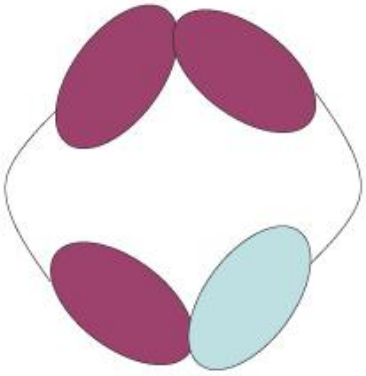
MLH1 – PMS1
wt

B



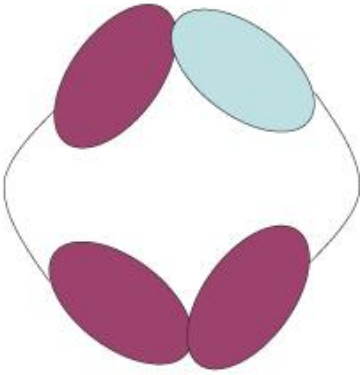
Double swap

C



MLH1 – MP
swap

D



MLH1 – PM
swap

Supplementary Table 5.1

Swapping MLH1-PMS1 N-terminal domains results in non-functional complexes displaying high mutator phenotypes

Genotype	*n*	Mutation Rate (10^{-7}), (95% C.I.)	Relative to wild type
Wild type	30	1.6 (1.4–2)	1
mlh1Δ	30	21300 (17500-25300)	13313
MLH1	10	22500 (12500-34500)	14063
ntd-mlh1-ctd-pms1, *MLH1* (MP swap)	20	22500 (17700-28400)	14063
ntd-pms1-ctd-mlh1, *MLH1* (PM swap)	10	17200 (12400-26000)	10750
ntd-mlh1-ctd-pms1, *ntd-pms1-ctd-mlh1*, (Both swaps)	10	27600 (15000-31300)	17250

The *mlh1* and *pms1* alleles listed were tested in the *lys2::insE-A*$_{14}$ mutator assay. For each swap allele the N-terminal domain (NTD) and C-terminal domain (CTD) subunit derivation is indicated. *Lys+* reversion rates were calculated in the strains EAY1365 (see Table 3.1). For each strain, the allele was expressed from an *ARS CEN* plasmid under the native promoter of the wild-type gene corresponding with the NTD after swap construction.